LA-UR-97–1366
April 15, 1997

# Model Uncertainty in Stochastic Simulation Models

**Contributions to the Development of a Statistical Framework for Joint Mobility Modeling**

**by**

**Michael D. McKay**
**Joanne R. Wendelberger**
**John D. Morrison**

**WORKING PAPER**

**Los Alamos National Laboratory**
**Los Alamos, New Mexico 87545–0600**

# MODEL UNCERTAINTY IN STOCHASTIC SIMULATION MODELS

*Abstract*

Model uncertainty in stochastic simulation models can be organized as a hierarchy of components, beginning at the top with structural model uncertainty and moving down through model parameter uncertainty, model input uncertainty and, finally, stochastic variability. This document investigates input and structural uncertainty for stochastic simulation models. It looks at methods for assessing the variability of model prediction due to choice of input values. For the case of discrete event simulation models, it examines how the contribution to overall prediction uncertainty from level of detail, aggregation, and alternative submodels might be assessed.

Key Words: model uncertainty, structural uncertainty, sensitivity analysis, uncertainty analysis, simulation modeling, analysis of computer experiments.

# Contents

## List of Figures



## List of Tables

## List of Equations

# 1 Introduction

The use of mathematical simulation models with “best estimate” or “expected” values for the simulation variables corresponds to a simplified conceptual view of reality. Such a view can easily be unrealistically precise in the face of known and expected real-world variability. The broader view of modeling acknowledges expected variability in reality and incorporates it into the simulation model. Such models, however, make stronger requirements on inputs and require more sophisticated methods for assessment than do the simplified models. In return, however, the models can provide more realistic predictions with higher quality assessments.

Uncertainty or variability in prediction for simulation models arises from several sources: mathematical or algorithmic structure of the model, values of the input variables, and random number streams. Characterizations of these sources of uncertainty enhance the value of model predictions by allowing for quantification of their precision, thus increasing the confidence one can have in them. This document develops and presents methods for assessing prediction uncertainty in simulation models with the objective of being able to provide additional means for decision makers to evaluate the adequacy and value of models and submodels. It uses a military mobility model as an example to demonstrate ideas and principles.

The approach to the development of methods for assessing structural model uncertainty in this document is to use specific model characteristics, as for a group of discrete event simulation models called “mobility models,” to produce a suitable context within which to formulate notions of model uncertainty, its origin and its measurement. Assessment of structural model uncertainty ideally involves comparisons of the predictions from several models, with the objective of observing how differences in model treatments or structures affect prediction. Formal methods therefore, usually assume that predictions from all relevant models can be computed implicitly or explicitly and combined some way into what is presumed to be a prediction better than that of any single model. Development of general and practical formal methods for assessing model uncertainty from these premises has evaded researchers, primarily for two reasons: difficulty in constructing and interpreting weights or probabilities associated with the validity[1] of constituent models, and difficulty in identifying suitably broad descriptions of alternative or competing models.

## 1.1 Scope

Previous work (McKay, 1995) is extended in several ways. For dealing with input uncertainty, the notion of “model” is expanded to allow for sampling variability of simulation variables. A more general variance decomposition is used for comparing input importance to simulation variability. The notion of importance as relating to statistical dependence is examined. Finally, a method to treat a structural uncertainty for a particular class of models is suggested.

[1] The term “validity” is used imprecisely as referring to correctness or truth.

## 1.2 Objectives

We address the needs of two groups: decision makers and model analysts. Decision makers and planners need the results obtained by model analysts who use quantitative statistical methods for evaluating models. We formulate a mathematical basis for model evaluation upon which we build statistical methodologies for answering the questions about model uncertainty. Properties of the methodologies are evaluated from a theoretical perspective so that the methods themselves may be evaluated.

Model evaluation methods described in this paper are intended to be a starting point for answering answer questions like:

- How do different input values affect model prediction?
- Which inputs drive output calculations?
- How might strong and weak characteristics of alternative models and submodels be identified?
- Are there ways to intelligently select among different modeling techniques for different parts of the modeling process?
- Can aspects of mobility models to carry forward to new models and to link with combat models be identified?
- Are there formal analysis methods for ranking alternative models?
- Where might research dollars be spent to advantage in improving data used for model inputs?

## 1.3 Overview

A mathematical model $m(\cdot)$ is as a formal statement of assumptions about a relationship between known quantities $x$ and unknown quantities $y$. The *structure* of the model defines an activity of determining characteristics of $y$ from those of $x$. The relation itself may depend on *parameters* $\theta$, which are generally treated as being distinct from $x$. When the values of $x$ are known to be not precisely accurate, we say that uncertainty in inputs $x$ propagates through the model $m(\cdot)$ to uncertainty in the output $y$. The same is true about uncertainty in parameter values $\theta$. These two types of uncertainty are referred to as *input uncertainty* and *parameter uncertainty*, respectively. Input and parameter uncertainty include sets of plausible values and, usually, probability distributions defined on the sets of values. The probability distributions are termed *epistemic* or *subjective*, indicating level of knowledge or belief of belief, or *aleatory* or *objective* indicating observational basis in objective reality. Because input and parameter uncertainty are assessed for individual models, it has been possible to develop generally acceptable methods of analysis, as in Helton (1993). and McKay (1995) General treatments of uncertainty about the structure of $m(\cdot)$, called both *structural model uncertainty* or just *model uncertainty*, have not been successful.

A treatment of model uncertainty which parallels the treatment of input uncertainty would address a model known to be not precisely accurate. Passing over questions of precision of numbers, models of physical systems are universally accepted to be approximations and, therefore, not completely accurate by construction! Therefore, one might try to finesse the question of degree of accuracy by formally representing a countable collection of models $\{m_i\}$ together with a probability distribution $P(m_i)$ which, somehow, describes the "correctness" of each model. (see, for example, Apostolakis, 1990, and Zio and Apostolakis, 1996, Laskey, 1996, and Draper, 1995,). Alternatively, one can view the predictions of each model in the collection as information, complete with prior and posterior (subjective) probability distributions suitable for Bayesian analysis. (Winkler, 1993 and Draper, 1995). In both cases, however, one must identify the set of alternative models and probability distributions. Fundamental difficulties associated with these tasks are identified by several authors, including Atwood (1993).

Putting aside the issue of probability distributions defined on a space of models, one can consider a family of plausible, alternative models defined implicitly. For example, one might formally construct a general family of models as realizations of a stochastic process super model, as do Sacks, Welch, Mitchell and Wynn (1989) in support of the analysis of computer experiments. Such models are common in the analysis of time series and spatial statistics. Applied in this study, the approach is to use of a suitable prototype model to provide a means to address reasonable questions regarding structural model uncertainty.

## 2 Prediction Uncertainty

We suppose that a decision maker or planner, presented with a model prediction, wants to know how close to it the outcome in reality can be expected. More informatively, the planner wants a range of plausible variation—a set of alternative predictions—about the model prediction and the likelihood that reality is be within the range. From the perspective of mobility models, the model prediction is a *plan* or, more precisely, a *realization of a plan* which is a set of events with completion times. An approach to examining the difference between prediction and reality is to treat the events as fixed completion times as variable. Letting $y$ denote the completion times from the model and $w$ the unknown actual (future) times, attention is directed to the "difference" between $y$ and $w$, which we denote by

$$\delta = \|y - w\| \,.$$

Two things the planner can ask are

- Analyst, please tell me what to expect about $\delta$.
- Analyst, please tell me which (input) variables to be careful about during plan execution in order to make the plan work as expected and to minimize $\delta$.

Because a plan comprises many events, it is practical and useful to summarize it by way of a simple *model observation* $y$ for purposes of comparison and evaluation. For mobility models, $y$ might be the completion times mentioned above, or Total Cargo or some other "Measure of Effectiveness." For practical reasons, $y$ is of low dimension, much less then the number of events in a plan.

A singularly complicating factor in this exercise is, try as one might, repeated attempts to execute exactly the same plan results in different execution times. We assume the reason is that in the execution of any plan there are events not under complete control. Therefore, the completion times, at least, are subject to an uncertainty designated as random or *stochastic* in nature. Whether or not the times are truly random is immaterial. What is important is we view them as having a stochastic element and we model them as random. As a result, the model observation $y$ is a random variable and we refer to reality $w$ as a stochastic process. Therefore, modeling and comparison between model prediction and reality must recognize stochasticity.

A planner or decision maker is much more interested in the probability distribution of plans or the probability distribution of the model observation $y$ rather than in any single plan or value of $y$. The planner would expect the probability distribution of model predictions $y$ to be close to that of $w$ in reality. Although the planner might ask that at least the expected value of $y$ from the model be an unbiased estimator of the expected value of $y$ in reality, such a condition alone is not usually a sufficient basis for accepting model predictions as valid. On the other hand, the real probability distribution of $w$ is unknown, so that even perfect knowledge of the probability distribution of the model prediction $y$ may not answer all questions about the validity of the model.

The comparison of model predictions $y$ with true values $y$ from reality is properly a part of model validation. As such, we only address it in passing in this paper. We are principally concerned with the characterization of probability distributions of the model prediction $y$, one of which we call the *simulation distribution*. The simulation distribution of $y$ is so important that we often think of it as the true prediction of the model. That is, we think of a simulation model as producing, at least implicitly, the set of values of $y$ together with the probability distribution of those values. It is understood that characterization of the simulation distribution is a necessary part of model validation and that the simulation distribution of $y$ is useful to the decision maker even without complete model validation.

The purpose of *uncertainty analysis* is to quantify and characterize uncertainty in the model prediction $y$ which, for stochastic simulation models, is an estimated characteristic of the simulation distribution. In previous work (McKay, 1995), $y$ was assumed to be determined without error. In this paper, however, estimation error is acknowledged as a necessary reality. Therefore, when investigating model uncertainty, estimation error becomes important.

A complete characterization of the simulation distribution of $y$ may be beyond reach, so that, in practice, an *expected value* solution might be used in lieu of an estimated probability distribution. Thus, $y$ might be an estimate of the expected value of $y$. We point this out because it happens and not to recommend it as a sufficient way for decision makers to use stochastic simulation models.

Figure 1 summarizes the are three sources of variation associated with uncertainty analysis of stochastic simulation models: structural uncertainty, input uncertainty, and simulation variability. We proceed to discuss these in the context of a mobility model. An abstract, mathematical discussion is delayed until the next section.

- Structural uncertainty arises from the selection of input variables and the relationships among them in the model.
- Input uncertainty refers to the values used for the inputs variables once they have been selected for use in a model.
- Simulation or sampling variability comes from the effect of random number streams in the simulation model.

Figure 1. Sources of variation in model uncertainty

For mobility models, the input variables characterize a scenario with initial conditions, define model components, and specify assumptions and rules governing the dynamics of processes in the simulation model. Inputs can be categorized as being *scenario descriptor variables* or *process variables*. Examples of scenario descriptor variables are ones specifying requirements, availabilities, numbers of carriers, and so forth. Examples of process variables are ones which characterize carriers as to capacity, landing requirements, and the like. Other process variables characterize the dynamics of the model by defining, for example, simulation probability distributions for loading times and travel times. Structural uncertainty arises from the selection of input variables and the relationships among them in the model. Input uncertainty refers to the values used for the inputs variables once they have been selected for use in a model. Because of the way we have chosen to relate structural and input uncertainty, input uncertainty is said to be "nested" in or "within" structural uncertainty.

Simulation variability in the model is the counterpart of *stochastic variability* or *stochastic uncertainty* in reality (Helton, 1994), and is associated with stochastic simulation models. Not all mobility models are stochastic and those that are may be used in a deterministic mode. In stochastic simulation mobility models, simulation variability comes from internal model variables, like service times, travel times and repair times, which behave like random variables within the model. These variables, which we call *stochastic variables* or *simulation variables*, are defined through random number streams within the simulation model. They are the reason for the simulation probability distributions of the model outputs. A consequence of sampling variability is estimation error, which is the difference between a population quantity and its sample estimate.

We are careful to distinguish between input uncertainty and simulation variability. Although each is characterized by a probability distribution, the two sources of variation are fundamentally different. Simulation variability is a representation of a variability expected in actuality. For example, we expect individual travel times for identical carriers identically loaded to be different. In the simulation model, we treat the simulation variable representing travel time as a random variable. We say that simulation variability is *objective variability*. On the other hand, input uncertainty often refers to one's state of knowledge: it tends to be, but is not necessarily, *subjective* rather than objective. Thus, one may use a (subjective) probability distribution to describe the available number of carriers of a certain type. At other times, however, uncertainty in an input value might be due to sampling or experimental error associated with estimation of the true input value. In those cases, the (objective) probability distribution describes the variability in a statistical process rather than one's degree of belief in a value. The distinction between subjective and objective variability is not always sharp and clear.

# 3 Components of Model Evaluation

Uncertainty analysis lies within a larger area of model evaluation, as outlined in Figure 2. An important objective of model evaluation is to demonstrate model credibility. The various components of model evaluation establish consistency and credibility of predictions during both model development and model application phases. Common terms used in discussion of model evaluation are:

- Diagnostic testing: fire it up and see what happens.
- Verification: testing for consistency between abstraction and software implementation (Bratley, et al., 1987, page 8)
- Validation: testing for consistency between model prediction and real world (data) (Bratley, et al., 1987, page 8)
- Uncertainty analysis: quantifying variability in prediction and identifying contributions to prediction uncertainty.

| **Model Evolution** | **Model Evaluation** |
|---|---|
| Real World | |
| ↓ | |
| Mathematical Abstraction | Logic validates abstraction and conceptual model. |
| ↓ | |
| Computational Model Construction and Software Implementation | Verification and testing to assure implementation is consistent with mathematical abstraction. |
| ↓ | |
| Calibration | Validation compares calibrated prediction with observation. |
| ↓ | |
| Numerical Calculation of Prediction | Statistical uncertainty analysis of numerical calculations for relevant scenarios. |
| ↓ | |
| Supplantation | Update or replace model. |
| ↓ | |
| Model Extinction | |

Figure 2. Life cycles in modeling

# 4 Formal Definitions for Simulation Models

## 4.1 Simulation Model

A *simulation model* is a mathematical model which describes a system by a sequence of states following from an initial state. It is represented by a triple $m = (\Omega, \mathbf{P}, T)$ in Figure 3. The state space $\Omega$ is finite or countably infinite (discrete).[2] The state of the system is given by the model at times in the index set $T$. The index set is either the set of nonnegative integers, for discrete time simulation, or else the nonnegative real numbers, for continuous time simulation.[3] The rules which determine transitions are encoded in the transition probability function $\mathbf{P}$. The word "probability" is used in the definition of $\mathbf{P}$ because its rules are allowed to contain stochastic or random factors, which is the usual case. The term "simulation model" is used synonymously with *stochastic simulation model* in this document.

$$
\begin{aligned}
\Omega &= \text{state space for a model} \\
\mathbf{P} &= \text{transition probability function for a model} \\
T &= \text{index set (of time points)}
\end{aligned}
$$

Figure 3. Components of a simulation model

In general, both the progression or sequence of states and the corresponding sequence of transition times are determined by the simulation model. In some cases, however, a simulation model might determine the sequence of times for a fixed and specified sequence of states.

## 4.2 Elements of a Simulation Model

### 4.2.1 Model Output

The *state of the system* as determined by the simulation model as a function of $t$ is denoted by

$$Y(t) \in \Omega \text{ for } t \in T \, .$$

Because $Y$ can be of high dimension, we usually use a transformation of it called a *model observation function,* defined on the state space $\Omega$. We denote the model observation function by

$$y(t) = \Gamma(Y(t)) \in \mathrm{R}^d$$

and, usually, assume that it is of low dimension, relative to the dimension of $\Omega$. Both $Y$ and $y$ are called model *outputs*.

[2] The extension to continuous state spaces is immediate.

[3] Most simulation models are of the finite-state-space, discrete-time variety.

In general, the output $Y(t)$ and observation function $y(t)$ are *random variables* by virtue of the stochastic element of the simulation model. Therefore, the output of a stochastic simulation model, viewed as a collection of random variables $\{Y(t), t \in T\}$, is a *stochastic process*. Whether or not $\{Y(t), t \in T\}$ or $\{y(t), t \in T\}$ are *Markov processes* depends on characteristics of the simulation model. Bailey (1964, page 137) says that the general queueing process in continuous time does not have the Markov character. However, there may be an observation function and a series of discrete time points at which $y$ is an embedded Markov chain.

### 4.2.2 Model Input

Model *inputs* or *input variables* are the *initial conditions* or initial state of the system, and indicated by

$$x = Y(0) \in \Omega$$

or

$$x = y(0) \in \mathrm{R}^d \, .$$

This notation is different from what is frequently used in that the distinction between model inputs and inputs values and model outputs and output values is blurred. While it seems to artificially complicate discussion for common deterministic models, it allows a general treatment of stochastic simulation models to include nonstochastic models and the usual engineering or prediction models.

### 4.2.3 Model Structure

Although *model structure* strictly means the triple $m = (\Omega, \mathbf{P}, T)$, it is common to use the term to mean the transition probability function $\mathbf{P}$. The transition probability function determines the value of $Y(t)$ from the state space $\Omega$ as follows.[4]

Let $\omega_i$ and $\omega_j$ be arbitrary states in a discrete state space $\Omega$, and $s < t \in T$. The elements of the transition probability function for discrete state spaces are the conditional probabilities that the system is in state $\omega_j$ at time $t$ given that it is in state $\omega_i$ at time $s$. That is,

$$\mathbf{P}_{ij}(s,t) = \Pr\{Y(t) = \omega_j \mid Y(s) = \omega_i\} \, .$$

For continuous state spaces, the elements of $\mathbf{P}$ are defined for times $s, t \in T$ and arbitrary state $\omega$ and (Borel) subset $E \subset \Omega$ by

$$\mathbf{P}_{E,\omega}(s,t) = \Pr(Y(t) \subset E \mid Y(s) = \omega) \, .$$

Detail or (fine) structure has to do with the smallest predictable subset $E$.

The probabilistic aspect of a simulation computation within $m(\cdot)$ is accounted for through pseudorandom number streams. The name “simulation variable” is used for a variable in the model determined directly from a random number. Simulation variables, which we denote by $z$, are directly related to the transition probability function $\mathbf{P}$.

[4] We are only treating discrete state spaces.

### 4.2.4 Model Parameters

The transition probability function can be viewed as a *process* or *activity model* of the mechanism governing changes from state to state. Activities typically depend of parameters that are independent of initial conditions and the initial state of the system. Let $\{\mathbf{P}_\theta$ or $\mathbf{P}(\theta),\ \theta \in \Theta\}$ denote a collection of transition probability functions indexed by $\theta$. We call $\theta$ *model parameters*. Model parameters often, though not necessarily, appear explicitly in functional forms as, for example, in

$$p_\theta(s,t) = e^{-\theta|t-s|}\ .$$

## 4.3 Example of a Simulation Model

The football game analogy. It is left as an exercise to the reader to complete this example.

$$\begin{aligned}
\omega = &\ \text{a 'freeze frame' of a football game}, \omega \in \Omega \\
Y(t) = &\ \text{the football game, minute by minute, } Y \in \Omega, \\
&\ \text{for } t = 1, 2, \cdots, 60 \\
y(t) = &\ \text{the score of the game } Y(t)\text{. Observation function on the space } \Omega \\
m = &\ \text{a 'simulation model,' which plays a football game} \\
&\ \text{and more, like } \mathbf{P}, \cdots
\end{aligned}$$

$$\begin{aligned}
\text{Domain} &= \text{Football} \\
\text{Observation} &= \text{Pitt vs KC}
\end{aligned}$$

## 4.4 Prediction Simulation Distribution

A natural description of the states $Y$ or observations $y$ that can be produced by the model $m(\cdot)$ with inputs $x$ is the probability space

$$(\Omega, \mathcal{A}, \mathcal{P}_x)$$

where $\Omega$ is the set of all possible states or observations on states, and $\mathcal{P}_x$ is a probability measure (distribution) defined on the elements of a $\sigma$-field $\mathcal{A}$ of subsets of $\Omega$. For simplicity, we let $\Omega$ be defined independently of any particular model but specific to the problem or scenario under study. The probability measure $\mathcal{P}_x$, on the other hand, depends on and is determined by the model structure and the values of the input variables. That is, the likelihood associated with any particular state depends on the model and input values used. In this sense, $\mathcal{P}_x$ is a *conditional* probability measure. The model observation $y$ is a random variable on $\Omega$. For discussion purposes, we treat $y$ as a one dimensional map of $\Omega$ into the real line $\mathrm{R}$. We define a right-continuous distribution function as

$$F_{y|x}(v \mid x) = \mathcal{P}_x(\{\omega \in \Omega : y(\omega) \leq v\})\ .$$

$F_{y|x}$ is called the *prediction simulation distribution function* or *simulation distribution* of $y$, induced by the simulation variables $z$ and conditional on $x$ and $m(\cdot)$. Usually, $F_{y|x}$ or some property of it like its mean or variance is the focus of investigation in analyses of model prediction uncertainty and relative importance of different sources of uncertainty.

### 4.5 Model Prediction

For deterministic models, a prediction is a real variable. For stochastic simulation models it is a random variable whose values are described by the simulation distribution $F_{y|x}$. A single output value $y$ from a stochastic simulation model is of little value when there is any significant variation due to simulation variables $z$. Therefore, simulation models are run several times with the same input values to produce a sample of output values, $y_1, y_2, \cdots, y_n$. A sample estimator $\tilde{y} = \tilde{y}(y_1, y_2, \cdots, y_n)$ is calculated from the sample values. Although $\tilde{y}$ might be a single realization of $y$, it is common to take it to be the simple average from a set of simulation runs, or a weighted average from an importance sample, or a sample estimator of the simulation distribution, $F_{y|x}$. Some common sample estimators are described in Table I, where population values refer to the simulation distribution, $F_{y|x}$.

TABLE I
Sample estimators for simulation models

| **Sample Estimator $\tilde{y}$** | **Population Value** |
|---|---|
| mean $\tilde{y} = \overline{y} = \frac{1}{n}\sum y_i$ | $\mu_y$ |
| standard deviation $\tilde{y} = \sqrt{\frac{1}{n-1}\sum (y_i - \overline{y})^2}$ | $\sigma_y$ |
| empirical cumulative distribution function (cdf) $\overline{y}(v) = \frac{1}{n}(\# y_i \leq v)$ | $F_y$ |
| quantiles of the empirical cdf $\tilde{y}_\alpha = v$ such that $\frac{1}{n}(\# y_i \leq v) = \alpha$ | $F_y^{-1}(\alpha)$ |
| maximum $\tilde{y} = \max_i \{y_i\}$ | $E\left(n^{th} \text{ order statistic}\right)$ |
| minimum $\tilde{y} = \min_i \{y_i\}$ | $E\left(1^{st} \text{ order statistic}\right)$ |

# 5 World, Reality and Model

## 5.1 View of Reality

Metaphorically speaking, reality[5] is created in the world by nature using the ultimate model in Figure 4, which provides the ultimate model against which simulation models are compared. A real system in some state $W^* = W^*(t)$ is acted upon by a complex process that causes it to change states. We assume that the evolution of the real system is described in a manner similar to that describing a simulation model, namely, that there is a state space $\Omega^* = \{\omega^*\}$ and a probability transition function $\mathbf{P}^*$, such that for two times $t > s$ and a discrete state space,

$$\mathbf{P}^*_{ij}(s,t) = \Pr\big(W^*(t) = \omega^*_j \mid W^*(s) = \omega^*_i\big) \ ,$$

For a continuous state space and two times $t > s$,

$$\mathbf{P}^*{}_{E^*,\omega^*}(s,t) = \Pr(W^*(t) \subset E^* \mid W^*(s) = \omega^*) \text{ for } E^* \subset \Omega^* \ .$$

$$\begin{aligned}
\Omega^* &= \text{state space for reality} \\
\mathbf{P}^* &= \text{transition probability function for reality} \\
T &= \text{index set (of time points)} \\
W^*(t) \in \Omega^* &= \text{state of reality at time } t \in T
\end{aligned}$$

Figure 4. Ultimate model describing reality

The ultimate model is hypothetical: the state vector $\omega^* \in \Omega^*$ is a hypothetical collection of characteristics and properties of a state, for which transitions are determined by a hypothetical transition function $\mathbf{P}^*$. The ultimate model (nature's) is unknown.

[5] Nature is who; reality is what; world is where; model is why.

$$\Omega \subset \Omega^* \text{ a subspace of the state space of reality}$$
$$\mathbf{P} = \text{a transition probability function on } \Omega$$
$$\omega \in \Omega \text{ the projection of } \omega^* \in \Omega^*$$
$$T = \text{index set}$$
$$W(t) \in \Omega \text{ the projection of } W^*(t) \in \Omega^*$$
$$w = \Gamma(W) = \text{observation function on reality}$$

Figure 5. Theory of reality

## 5.2 View of a Model

For a modeler, knowledge of reality is partial and incomplete, and a theory like Figure 5 is used to explain it. A simulation model $m(\cdot)$ incorporates the theory of reality as $m = (\Omega, \mathbf{P}, T)$. The idea that the model $m(\cdot)$ considers much less detail in the modeled system than actually exists in nature is reflected in the dimension of $\Omega$ being much less than that of $\Omega^*$. Nevertheless, we assume that in some limiting sense, $\Omega \subset \Omega^*$. That is, that the representation of states within the model is by a subset of its characteristics in nature, or that any idealized quantities used as characteristics of the state of the system have a physical counterpart, at least approximately.

A complete understanding of state transition mechanisms in nature requires knowledge of both the transition probability function $\mathbf{P}^*$ and of the state space $\Omega^*$. For the simulation model, it is less reasonable to assume that $\mathbf{P} : \Omega \to \Omega$ is $\mathbf{P}^*$ restricted to $\Omega$ than to assume $\Omega \subset \Omega^*$. Nevertheless, we assume that in some limiting sense, $\mathbf{P}$ can be made to approach $\mathbf{P}^*$. Therefore, we assume that there is no fundamental reason that simulation models cannot be of the same quality as nature's ultimate model.

## 5.3 Prediction

Letting $W$ denote the unknown state of reality and $w$ the observation function on reality, a model prediction[6] might be of

- the instance itself, $w$
- its expected value $E(w)$
- its complete probability distribution$f_w(w)$.

[6] Model prediction are usually based in input values $x$. Therefore, of importance are predictions conditional on the value of those inputs: the conditional instance itself, $w \mid x$, its conditional expected value $E(w \mid x)$, and so forth. Moreover, given that $x$ is not precisely known, it is important to quantify variability of these quantities relative to a probability distribution on $x$. That is, it is important to have results from an analysis of input uncertainty.

The quality of prediction is associated with an equation that relates model prediction $y$ and reality $w$, such as

$$w = y + (\text{simulation variability}) + e\ .$$

Simulation variability, a inherent property of the model, arises from choice of the random number stream. The prediction error $e$, on the other hand, is controllable through use of existing information (and more accurate modeling through $\mathbf{P}$) and through use of additional information (and modeling more detail in $\Omega$).

The prediction error has three components: error due to model structure, error due model parameters conditional on structure, and error due to input values conditional on structure and parameter values. The problem we address is quantification of changes prediction error due to alternative structures.

We assume that errors due to model structure can be reduced by increased model resolution or model refinement. Refinement means increasing the level of detail in the description of a process activity or in the description of model objects (actors) and states. However, if the limitations of modeling are not realized, refinement can easily introduce bias (inaccuracy) while reducing variability of prediction, as illustrated in Figure 6. Additionally, for a fixed or controlled level of bias, the cost of achieving reduced variability might go as in Figure 7.

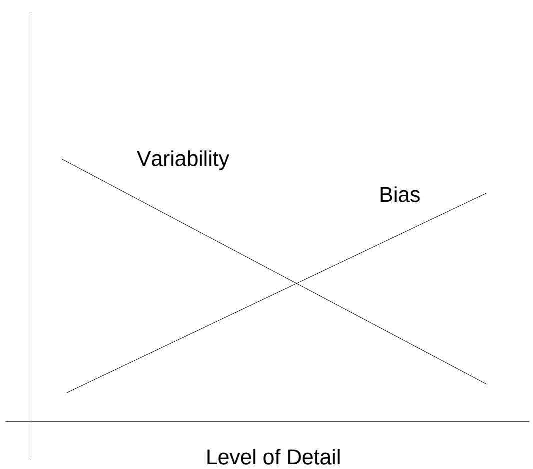


Figure 6. Possible trade-off between variability and bias

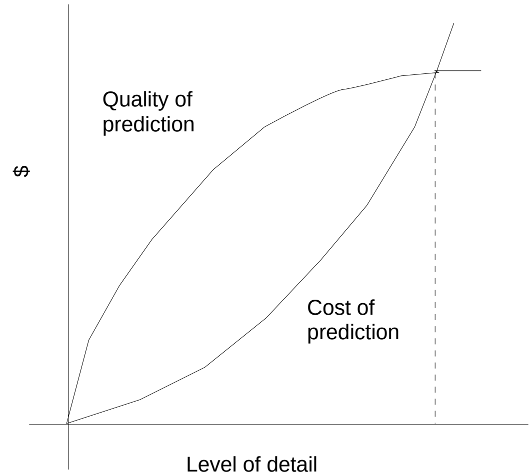


Figure 7. Diminishing returns with increasing costs

# 6 Model Uncertainty in Stochastic Simulation

Uncertainty is introduced into a simulation model in several ways, and it is manifest as uncertainty or variability in the model output $y$. Four sources of uncertainty, or variation, are associated with a stochastic simulation model: structural uncertainty, parameter uncertainty, input uncertainty, and simulation variability. Taken together, they constitute the sources of variation for *model uncertainty*.

Model uncertainty, summarized in Figure 8, refers to the variability in outputs due to sampling of random numbers, plausible alternative input values, plausible alternative parameter values, and plausible alternative model structures. We denote these as simulation or sampling variability, input uncertainty, model parameter uncertainty, and structural model uncertainty, respectively. Model uncertainty has a *nested structure* or hierarchy:

- Model $m(\cdot)$. Structural or model uncertainty comes from plausible alternative choices of $m(\cdot)$. A choice of $m(\cdot)$ includes the designation of the parameter and input variable sets but not necessarily their values.
  - Parameter values $\theta$ used in the model. Parameter uncertainty comes from plausible alternative choices of values $\theta$.
    - Input variable values $x$ used in the model. Input uncertainty comes from plausible alternative choices of values $x$ for the fixed model $m(\cdot)$ and parameter values $\theta$.
      - Simulation variable values $z$ used in the model. Simulation variability given $m(\cdot)$, $\theta$ and $x$ comes from sampling variability induced by the use random numbers in the representation of stochastic events and phenomena in the model.

**Model Uncertainty** refers to the variability of model output values due to sampling of random numbers (simulation variability), plausible alternative input values (input uncertainty), plausible alternative parameter values (parameter uncertainty), and plausible alternative model structures (structural uncertainty).

Figure 8. Model uncertainty

### 6.1 Structural Model Uncertainty

The highest level of model uncertainty is structural uncertainty. Structural model uncertainty refers to variability in model prediction arising from plausible alternative model structures. Structural uncertainty is the most difficult source of variation to study, in general, because of the difficulty of defining the probability space of alternative models. Because of this, most uncertainty analyses incorporate structural uncertainty in only limited ways, if at all. Structural uncertainty is characterized in Figure 9.

**Structural Uncertainty** refers to variability in model prediction $y$ arising from plausible alternative model structures.

$\mathcal{M}$: space of plausible models, $m \in \mathcal{M}$
$x$: input value, $x \in D$
$f_{z|x}$ : conditional probability distribution of simulation variables
$g_m$: probability function on $\mathcal{M}$
$g_y$: probability function distribution of $y$ induced by $\left(f_{z|x}, x, g_m, \mathcal{M}\right)$

Figure 9. Structural uncertainty

### 6.2 Model Parameter Uncertainty

It sometimes happens that the family of alternative models can be conveniently defined parametrically. In such cases, model parameter uncertainty refers to variability in model prediction arising from alternative choices of the value of a parameter $\theta$. In this document, parameter uncertainty is just a higher level of input uncertainty, which is defined in the next section.

### 6.3 Model Input Uncertainty

Input uncertainty refers to variability in model output due to plausible alternative input values. Input uncertainty reflects our view about the likelihood of appropriate values $x$ to use in the simulation model. Input uncertainty and simulation variability both contribute to the variability in $y$. In order to examine input uncertainty separately from simulation variability, it is convenient to view input uncertainty with respect to the simulation distribution of the model prediction. For example, we study the effect of input values on the mean of the simulation probability distribution, $F_{y|x}$, induced by the random number stream. Input uncertainty is characterized in Figure 10.

**Input Uncertainty** refers to variability in model prediction $y$ arising from plausible alternative input values for a given model structure.

$m(\cdot)$: model structure
$f_x$: probability distribution of input variables
$f_{z|x}$ : conditional probability distribution of simulation variables
$F_y$: probability distribution of $y$ induced by $\left(f_{z|x}, f_x, m(\cdot)\right)$

Figure 10. Input uncertainty

### 6.4 Simulation Variability

Simulation variability is fundamental to stochastic simulation models. It is the variability of prediction $y$ associated with different random number streams. The simulation variability of $y$ is characterized by the simulation probability distribution $F_{y|x}$. This distribution is a conditional probability distribution that depends on input values $x$ and the model $m(\cdot)$. Because $y$ is a random variable by design in simulation models, values of its population characteristics are usually estimated from a sample of values. The estimates of the mean or expected value of $y$ and its cumulative distribution function are often treated the output of a simulation model. Simulation variability is characterized in Figure 11.

**Simulation Variability** is the variably in $y$ arising from random number streams.

$m(\cdot)$: model structure
$x$: input value, $x \in D$
$f_{z|x}$: conditional probability distribution of simulation variables
$F_{y|x}$: probability distribution of $y$ induced by $\left(f_{z|x}, x, m(\cdot)\right)$

Figure 11. Simulation variability

# 7 Prediction Uncertainty in Stochastic Simulation

The counterpart of $y$, $w$ is observed from a real-world mobility. We say $y$ is a “realization” from the model and $w$ is “realization” from reality. Questions to answer are:

- How close to $w$ can $y$ be expected; what is the magnitude of $\|y - w\|$?
- How close is the probability distribution $F_{y|x}$ of $y$ to $F_w$ of $w$? What is their joint probability distribution?
- In what ways do different input values and simulation variables affect $\|y - w\|$?

Each question above depends on the value of $x$. Therefore, the way in which the input values $x$ are treated determines how the questions are answered. If $x$ is assumed known, it is appropriate to use the probability distribution of $y$ conditional on $x$. In that event, the simulation distribution of $y$ contains all of the information from the model prediction. However, it is unlikely the values of $x$ would be known with certainty. That is, input uncertainty as characterized in Figure 10 must be considered. With input uncertainty present, it is correct to use the unconditional distribution of $y$, which we call the *prediction distribution*. The prediction distribution induced by both the input random variables $x$ and the simulation random variables $z$ characterizes prediction uncertainty in $y$. Although the model observation $y$ has two meaningful probability distributions, the one most relevant to prediction uncertainty is its unconditional distribution.

The prediction distribution of $y$ is the (marginal) distribution induced by $x$ and $z$,

$$
\begin{aligned}
y &\sim F_y(v; m) \\
F_y(v; m) &= \text{Pr}\{y \leq v \mid \text{model } m\} \\
E(y) &= \mu_y \\
\text{Var}[y] &= \sigma_y^2
\end{aligned}
$$

It is the model counterpart of the distribution $F_w$ of $w$.

We assume that the probability distribution of $x$ is appropriate to the scenario or situation being modeled, so that $y$, as a function of the random variable $x$, is an appropriate model counterpart to $w$. The sensitivity of the uncertainty analysis of $y$ to changes in the input distribution of $x$ may be studied using techniques suggested by Beckman and McKay (1987).

When comparing two random variables, it is common to use the mean (expected) squared difference as a measure of the difference. The mean squared difference for the prediction $y$ and $w$ is

$$
\begin{aligned}
\|y - w\|^2 &= E(y - w)^2 \\
&= (\mu_y - \mu_w)^2 + E(y - \mu_y)^2 + E(w - \mu_w)^2 \\
&= (\mu_y - \mu_w)^2 + \sigma_y^2 + \sigma_w^2 \,.
\end{aligned}
\tag{7.1}
$$

In the mean square sense, $\|y - w\|$ depends on

- bias, the difference between the expected values of $y$ and $w$,

- variability of $y$ about its expected value, and
- variability of $w$ about its expected value.

We use a sample $y_1, y_2, \cdots, y_n$ of $y$ to construct a sample mean $\tilde{y} = \frac{1}{n}\sum y_i$, say, which also has the same expected value $\mu_{\tilde{y}} = \mu_y$ but a smaller variance. Therefore, in this paper and without loss of generality, we focus on $y$ as the model prediction, realizing that it often represents a sample mean value. The probability distribution of a sample estimator $y$ depends on whether the sample is a random sample, a Latin hypercube sample, and importance sample, and so forth.

As discussed in Section 6, model uncertainty (MU) has three components: structural uncertainty (SU), input uncertainty (IU) within structural uncertainty, and simulation variability (SV) within structural uncertainty and input uncertainty. The term "within" is a way of addressing the conditional nature of the hierarchy. Model uncertainty refers to the variability in outputs due to sampling of random numbers (simulation or sampling variability), plausible alternative input values (input uncertainty), and plausible alternative model structures (structural uncertainty). Symbolically, we write

$$\mathrm{MU} \approx \mathrm{SU} \oplus (\mathrm{IU} \mid \mathrm{SU}) \oplus (\mathrm{SV} \mid \mathrm{IU, SU}), \tag{7.2}$$

by which we mean that the probability distribution representing model uncertainty is the product of a marginal contribution from structural uncertainty, a conditional contribution from input uncertainty conditional on structural uncertainty, and a conditional contribution from simulation variability conditional on both structural uncertainty and input uncertainty. We call the total combined contribution from input uncertainty and simulation variability *prediction uncertainty* (PU), conditioned on or for a given model structure. Thus, we can also write model uncertainty in terms of two components

$$\mathrm{MU} \approx \mathrm{SU} \oplus (\mathrm{PU} \mid \mathrm{SU}) \quad .$$

We say that prediction uncertainty itself has two components: input uncertainty and simulation variability within input uncertainty. Again, symbolically, we write

$$\mathrm{PU} \approx \mathrm{IU} \oplus (\mathrm{SV} \mid \mathrm{IU}) \quad .$$

We treat prediction uncertainty for one model structure at a time and not simultaneously for all structures in general. The probability distribution associated with prediction uncertainty is called the prediction distribution. Therefore, when we talk about prediction uncertainty, we do so in the context of a given model and, usually, not averaged over a set of models.

For deterministic models, prediction uncertainty has only one component, input uncertainty. For stochastic simulation models, the addition of simulation variability means that additional steps are needed in uncertainty analysis.

## 7.2 Errors in Prediction

Uncertainty analysis examines “errors” in model predictions. Error in prediction for model $m(\cdot)$ for a sample estimator $y$, whether it be the sample mean or standard deviation, for example, is the difference between $y$ and its prediction target $\tau$. The prediction target might be the expected value of $w$, $\mu_w$, or the distribution of $W$, $F_w$. It is common to examine the error in prediction $y - \tau$ in mean square. The difference between prediction and target expressed in the mean squared error sense is

$$\begin{aligned} \|y - \tau\|^2 &= E_x E_{z|x}\left[(y-\tau)^2\right] \\ &= E_x E_{z|x}\left[\left(y - \mu_{y|x} + \mu_{y|x} - \mu_y + \mu_y - \tau\right)^2\right] \\ &= E_x E_{z|x}\left[\left(y - \mu_{y|x}\right)^2\right] + E_x\left(\mu_{y|x} - \mu_y\right)^2 + E_x(\mu_y - \tau)^2 \\ &= E_x(\mathrm{Var}[y \mid x]) + \mathrm{Var}\left[\mu_{y|x}\right] + (\mu_y - \tau)^2 \\ &= E_x\left(\sigma^2_{y|x}\right) + \sigma^2_{\mu_{y|x}} + (\mu_y - \tau)^2 \,. \end{aligned} \tag{7.3}$$

For the most part, the terms in the last line of Eq. 7.3 correspond, respectively, to the three fundamental sources of variation model uncertainty: (1) the random variable stream used in simulation runs, (2) values of the inputs that define model processes and modeled events, and (3) model structure, including specification of input variables and relationships among them. We use the terms “simulation error” or “estimation error” in reference to the difference

$$y - \mu_{y|x} \,,$$

“prediction error” in reference to

$$\mu_{y|x} - \mu_y \,,$$

and “prediction bias” in reference to the difference

$$\mu_y - \tau \,.$$

For deterministic models, $\mu_{y|x}$ is just $y$, so that the simulation error is zero. These terms and their variances are the basis for variance-based importance measures described in the Section 8, which follows from the discussion of importance in the next section.

We mention in passing that, sometimes, simulation variables $z$ are set to their expected values and the resulting output is used at the estimate of the simulation mean $\mu_{y|x}$. The result is called the “expected value solution.” The practice does not necessarily yield the intended results, namely, the expected prediction. The reason is that, in general,

$$E_{z|x}(m(x,z)) \neq m(x, E(z \mid x)) \,.$$

Moreover, the simulation expected value alone may be a completely inadequate characterization of the simulation distribution in as much as it says nothing about statistical variation about the mean.

### 7.3 Importance

The dependence of the prediction distribution on input uncertainty and simulation variability can be seen in the prediction distribution $y$. We examine simulation variability first, followed by input uncertainty.

The model output $y$ depends on simulation variables $z$ and input $x$. To simplify presentation, we suppose that $y$ is continuous and has a probability density function

$$f_y(t) = \frac{dF_y(v)}{dv} \,,$$

and write it as

$$f_y(v) = \int f_{y|x}(v) f_x(x) dx \,, \tag{7.4}$$

which shows that the prediction distribution of $Y$ is the weighted average of its simulation distributions, $\left\{f_{y|x}\,,\ x \in D\right\}$.

In a similar manner, for an arbitrary partition of $p$ inputs

$$x = x^p$$

into subsets $x^s$ and $x^{p-s}$, the prediction distribution is

$$f_y(v) = \int f_{y|x^s}(v \mid x^s) f_{x^s}(x^s) dx^s \,. \tag{7.5}$$

We assume that the importance of the inputs $x$ relative to the simulation variables $z$ is captured in the difference among the distributions in the family of simulation distributions $\left\{f_{y|x}\,,\ x \in D\right\}$ from Eq. 7.4. Similarly, the importance of the input subset $x^s$ is captured in the family of conditional prediction distributions $\left\{f_{y|x^s}\,,\ x \in D\right\}$ from Eq. 7.5.

Intuitively, $x$ is important relative to $z$ if different values of $x$ produce sufficiently different simulation distributions. In other words, $x$ is important if the distributions in $\left\{f_{y|x}\,,\ x \in D\right\}$ are sufficiently different from their (weighted) average value, the prediction distribution $f_y$. Similarly, the subset of inputs $x^s$ is important relative to the entire set of inputs if different values of $x^s$ produce sufficiently different condition prediction distributions in the family $\left\{f_{y|x}\,,\ x \in V\right\}$.

The comparison among functions in Eqs. 7.4 and 7.5 is a more general form of the problem of assessing the distance between two functions using, for example, Hellinger or Matusita's distance (see Kotz and Johnson, 1982, page 334). Other measures of distance between functions are the Kullback-Leibler distance and information number (Kullback, 1968), which are entropy-related measures. Before investigating these topics, we look at a simpler approach which has a proven worth.

# 8 Prediction Variance and Importance Indicators

## 8.1 Decomposition of the Prediction Distribution

We continue with the mathematical development of Section 7.3 for a "computer experiment" from which uncertainty and importance are to be studied. We denote general sample values by

$$y_1, y_2, \cdots, y_n$$

from which we compute the sample estimator

$$y = y(y_1, y_2, \cdots y_n)\ .$$

As before, $y$ denotes a model prediction from a sample of model runs. From Table I, for example, $y$ might be the mean of a sample of size $n$ of the model observation $Y$. For simple random samples, the values are independent and identically distributed. For other kinds of samples, this is not the case. By definition, $y$ is an *exchangeable* function of $y_1, y_2, \cdots, y_n$, meaning that the order of its arguments does not effect its value. To simplify presentation, we denote the probability density function of $y$ by

$$f_y(t) = \frac{dF_y(v)}{dv}\ .$$

For now, let the sampling schemes be simple random, where each sample value (replicate) $y_i$ corresponds to an independently sampled value of simulation variables $z_i$. In summary,

$$\begin{gathered} x \sim f_x(x),\ x \in \mathrm{R}^p \\ z_i \sim f_{z|x}(z) \text{ independent each } i = 1, 2, \cdots, n \\ y_i = m(x, z_i) \\ y = y(y_1, \cdots, y_n) \\ y \sim f_y(y),\ y \in \mathrm{R}^d\ . \end{gathered}$$

We quantify the uncertainty and assess importance of input subset through estimates of the prediction distribution $f_y$ and various decomposition thereof, which we now discuss.

In Section 7.3, we saw that the prediction distribution $F_Y$ of the model observation $Y$ could be written as the (weighted) average of its (conditional) simulation distributions. Similarly, the prediction distribution of the sample estimator $y$ can be written as

$$\begin{aligned} f_y(y) &= \int f_{y|x}(y) f_x(x) dx \\ &= \int \int f_{y,z|x}(y, z) f_{z|x}(z) f_x(x) dz dx\ . \end{aligned}$$

The joint conditional distribution of $y$ and $z$ conditional on $x$ is a degenerate distribution because $y$ is a function of $z$, uniquely defined by it given $x$. However, that is not true in general when conditioning is on a subset $x^s$ of inputs. In that case, the relevant distributions can be written as

$$\begin{aligned} f_y(y) &= \int f_{y|x^s}(y) f_{x^s}(x^s) dx^s \\ f_{y|x^s}(y) &= \int f_{y,z|x^s}(y,z) f_{z|x^s}(z) dz \ . \end{aligned} \tag{8.1}$$

These last two equations are a foundation for assessment of importance and construction of importance indicators. As argued before, the notion of importance, in this case for the subset $x^s$, is captured in the dissimilarity within the family

$$\left\{ f_{y|x^s}, \ x \in D \right\} ,$$

which shows the effect of different values of $x^s$ on $y$. Strictly speaking, we are defining "importance" through statistical dependence: if $x^s$ and $y$ are statistically independent, we say that $x^s$ is not at all important. Under independence,

$$f_{y|x^s}(y) = f_y(y) \text{ for all values of } x^s \ .$$

At the other extreme of independence,

$$f_{y|x^s}(y) = \text{ point mass function at the value } y(x^s).$$

This perspective allows a wide range of approaches to importance. In the next section we investigate using variance.

## 8.2 Decomposition of Prediction Variance

The mean value and variance of the prediction probability distribution $f_y$ are fundamental attributes commonly used as proxies for the full distribution, even though it is only in special cases, like that of the normal distribution, that the mean and variance uniquely identify the distribution. The mean and variance often contain enough information to suffice for analysis and comparison of $f_y$ to the family of distributions $\left\{ f_{y|x^s} \right\}$. Prediction mean and variance are given by

$$E(y) = \int y f_y(y) dy = \mu_y$$

and

$$\begin{aligned} \mathrm{Var}[y] &= E(y - \mu_y)^2 \\ &= \int (y - \mu_y)^2 f_y(y) dy \\ &= \sigma_y^2. \end{aligned}$$

They represent two aspects of the difference between the prediction distribution and a conditional prediction distribution,

$$\left\| f_y(y) - f_{y|x^s}(y \mid s_x) \right\| , \tag{8.2}$$

which measures the importance of the input subset $x^s$.

We can write the prediction variance $\sigma_y^2$ of $y$ in terms of the conditional prediction variances of the distributions $f_{y|s_x}$ from Eq. 8.1 as follows. The notation, though very cluttered, is needed to make explicit the roles of the various random variables.

$$
\begin{aligned}
\sigma_y^2 &= \mathrm{Var}[y] \\
&= \mathrm{Var}_{x^s}\left[E_{x^{p-s}|x^s}(y \mid x^s)\right] + E_{x^s}\left[\mathrm{Var}_{x^{p-s}|x^s}(y \mid x^s)\right] \\
&= \mathrm{Var}_{x^s}\left[E_{x^{p-s}|x^s}(y \mid x^s)\right] + \\
&\quad E_{x^s}\left\{\mathrm{Var}_{z|x^s}\left[E_{x^{p-s}|x^s,z}(y \mid x^s, z)\right] + E_{z|x^s}\left[\mathrm{Var}_{x^{p-s}|x^s,z}(y \mid x^s, z)\right]\right\} \\
&= \mathrm{VCE}(y \mid x^s) + E[\mathrm{CVCE}(y \mid z; x^s)] + \overline{\sigma}_e^2
\end{aligned}
\tag{8.3}
$$

This equation shows the decomposition of the prediction variance from which importance indicators for stochastic simulation models are constructed in Section 8.3.

The three terms in the last line of Eq. 8.3 derive from input uncertainty and simulation variability, and sum to a measure of prediction uncertainty. Using a simpler notation, we name them:

- *Variance of the Conditional Expectation (VCE) of* $y$, conditioned on the input subset $x^s$ is

$$
\begin{aligned}
\mathrm{VCE}(y \mid x^s) &= \mathrm{Var}_{x^s}\left[E_{x^{p-s}|x^s}(y \mid x^s)\right] \\
&= \int \left(y - \mu_{y|s_x}\right)^2 f_{y|s_x}(y)dy
\end{aligned}
$$

- *Conditional Variance of the Conditional Expectation (CVCE) of* $y$, conditioned on the input subset $x^s$ and simulation variables $z$ is

$$
\begin{aligned}
\mathrm{CVCE}(y \mid z; x^s) &= \mathrm{Var}_{z|x^s}\left[E_{x^{p-s}|x^s,z}(y \mid x^s, z)\right] \\
&= \int\int \left(\mu_{y|z,s_x} - \mu_{y|s_x}\right)^2 f_{y,z|s_x}(y,z)dydz \;;
\end{aligned}
$$

- *Partial Variance of the Conditional Expectation (PVCE) of* $y$ conditioned on the input subset $x^s$ is the expected value of the $\mathrm{CVCE}(y \mid z; S_x)$ over $x^s$,

$$
\begin{aligned}
\mathrm{PVCE}(y \mid z; S_x) &= E_{x^s}\left\{\mathrm{Var}_{z|x^s}\left[E_{x^{p-s}|x^s,z}(y \mid x^s, z)\right]\right\} \\
&= \int\int\int \left(\mu_{y|z,s_x} - \mu_{y|s_x}\right)^2 f_{y,z|s_x}(y,z) f_{s_x}(s_x)dydzds_x \,.
\end{aligned}
$$

The last term in Eq. 8.3 is the variance contribution due to the other inputs, $x^{p-s}$, and their interaction with the simulation variables $z$. In summary, then,

$$
\mathrm{Var}[y] = \mathrm{VCE}(y \mid x^s) + \mathrm{PVCE}(y \mid z; x^s) + \overline{\sigma}_e^2 \,. \tag{8.4}
$$

## 8.3 The Correlation Ratio

The amount of variability "explained" by another random variable $x$ is measured by the size of the variance of the expected value of $y$ conditioned on $x$ relative to the unconditional variance of $y$. This quantity is called the *correlation ratio*. As a measure influence, both the correlation ratio and its counterpart for the linear model, the correlation coefficient, were used by Pearson (1903) at the turn of this century. A good development of both statistics can be found in Kendall and Stuart (1979). Other references that provide development and applications to computer models are Krzykacz (1990), Iman and Hora (1990), Saltelli, Andres, and Homma (1993) and McKay (1995).

From Eq. 8.4, the prediction variance $\sigma_y^2$ is the sum of three components corresponding to an input subset $x^s$, simulation variables $z$, and a residual component. The importance of the input subset is measured by the ratio of its contribution to the total prediction variance. the ratio is called the correlation ratio,

$$\eta^2_{x^s} = \frac{\text{VCE}(y \mid x^s)}{\sigma_y^2} \; .$$

The importance of the simulation variables, given the subset $x^s$, is measured by the *partial correlation ratio*

$$\eta^2_{z.x} = \frac{\text{PVCE}(y; z, x^s)}{\sigma_y^2} \; .$$

Estimates for these quantities are given in Section 9.

## 8.4 Analysis Models

Importance indicators in some sense indicate the viability or strength of a variable subset as a predictor. The validity of the indicators depends on the assumptions under which they function. In particular, the correlation ratio becomes the correlation coefficient for subsets of size 1 $x$ (or the multiple correlation coefficient, in general) under a *linear analysis model*. As seen in Section 9, there are trade-offs between analysis assumptions and required sample sizes.

### 8.4.1 Linear analysis model

Under the linear analysis model, the expectation of $y$ conditional on an input subset $x^s$ is assumed to be approximately a linear function of $x^s$, Using the notation of Figure 12, the variance of the conditional expectation of $y$ is given by

$$\text{Var}[E(y \mid x^s)] = \beta^t \Sigma \beta \; .$$

Furthermore, under the usual assumption that the covariance matrix is of the form

$$\Sigma = \sigma^2 I \; ,$$

the problem of estimation of the conditional variance becomes a regression problem of estimating $\beta$ and $\sigma^2$.

### 8.4.2 General analysis model

The general analysis model makes no particular assumptions about the form of the expected value of $y$ conditioned on $x$, as indicated in Figure 13. Under the general analysis model, problem of estimation of the conditional variance becomes a problem in the estimation of variance components, analogous to random effects analysis of variance computations (see McKay, 1995).

For any arbitrary subset $x^s$ of inputs, the *linear analysis model* is

$$\begin{aligned} y &= E(y \mid x^s) + e \\ &= x^s\beta + e \end{aligned}$$

with $\beta$ as unknown fixed constants,

$$E(e) = 0$$

and

$$\text{Cov}[E(y \mid x^s), e] = \Sigma \ .$$

Figure 12. Linear analysis model

For any arbitrary subset $S_x$ of inputs, the *general analysis model* is

$$y = E(y \mid x^s) + e$$

with

$$E(e) = 0$$

and

$$\text{Cov}[E(y \mid x^s), e] = \Sigma \ .$$

Figure 13. General analysis model

# 9 Estimation of Importance Indicators

We perform an uncertainty analysis in a sequential manner: generating a two- or three-stage sample of model predictions $y$, examining results, forming input subsets, and repeating the process. The procedure is much less complicated than it might sound. Steps are described in detail in McKay (1995, pages 31–33). We present here an outline with extensions to stochastic simulation models.

## 9.1 One Sample

Let

$$\{x_i,\ i = 1, \cdots, n\}$$

be a simple random sample of inputs $x$ of size $n$ from $f_x$. In practice, we would likely use a Latin hypercube sample (McKay, Conover and Beckman, 1979). However, the expected value derivations below are easier for simple random samples. For each $x_i$, let

$$\{z_{ij},\ j = 1, \cdots, k\}$$

be a simple random sample of simulation variables $z$, independent for each $i$, from $f_{z|x}$. Let the resulting model predictions and "computer experiment" be

$$\begin{gathered}\{y_{ij},\ i = 1, \cdots, n; \quad j = 1, \cdots, k\} \\ y_{ij} = m(x_i, z_{ij}) \\ x_i \sim f_x \text{ iid} \\ z_{ij} \sim f_{z|x_i} \text{ conditionally iid.}\end{gathered}$$

Independent and identically distributed random variables are indicated by "iid."

The three sample means of values corresponding to the samples of simulation variables for each input vector value, grand mean, and input variable means are, respectively,

$$\overline{y}_i = \frac{1}{k}\sum_{j=1}^{k} y_{ij}\ , \quad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} \overline{y}_i\ , \quad \text{and } \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\ .$$

The correlation ratio and the correlation coefficient are estimated (with bias) from a sample of values following the analysis of variance decomposition of sums of squares in Figure 14. Estimates of the correlation ratio and the square of the correlation coefficient are indicated at the bottom of the figure. An advantage of using the estimators indicated is that they have the property of their population counterparts that $0 \le \widehat{\rho}^2 \le \widehat{\eta}^2 \le 1$, as implied in the figure. A disadvantage is that bias is introduced because of the covariance structure of the $y_{ij}$ induced by the sample design. For example, the expected value of the total sum of squares is, approximately,

$$\begin{aligned}E(\text{SST}) &= k(n-1)\text{Var}[y] + (k-1)E(\text{Var}[y \mid x]) \\ &= k(n-1)\sigma_y^2 + (k-1)\overline{\sigma}_e^2\ ,\end{aligned}$$

which can be driven to $nk\sigma_y^2$ with $n$ for fixed $k$. Similarly,

$$\begin{aligned} E(\text{SSB}) &= k(n-1)\text{Var}[E(y \mid x)] + (n-1)E(\text{Var}[y \mid x]) \\ &= k(n-1)\text{Var}[E(y \mid x)] + (n-1)\overline{\sigma}_e^2 \, , \end{aligned}$$

and

$$\begin{aligned} E(\text{SSW}) &= n(k-1)E(\text{Var}[y \mid x]) \\ &= n(k-1)\overline{\sigma}_e^2 \, . \end{aligned}$$

The ratio of expectations is given by

$$\frac{E(\text{SSB})}{E(\text{SST})} = \frac{(n-1)(k-1)\eta^2 + (n-1)}{kn - 1 - (k-1)\eta^2} \, .$$

In the limit with $n$,

$$\lim_{n\to\infty} \frac{E(\text{SSB})}{E(\text{SST})} = \left(1 - \frac{1}{k}\right)\eta^2 + \frac{1}{k}$$

The products of this step are estimates of the prediction distribution by way of sample means, variances, and empirical cumulative distribution functions, and estimates $\widehat{\eta}_x^2$ and $\widehat{\eta}_{z.x}^2$ of the correlation ratios reflecting the relative contributions to prediction variance from input variables and simulation variables, respectively. Figure 14 also indicates how one tests the adequacy of the linear analysis model and the linear regression approximation to the model. The strength of linear fit is measured by the ordinary (multiple) correlation coefficient $\widehat{\rho}_x^2$, computed for a single, $p = 1$, input in the figure for simplicity of presentation. The significance of the fit in the usual analysis-of-variance sense is tested with the F statistic,

$$F = \frac{\text{SSR}/p(=1)}{\text{SSW}/n(k-1)}$$

### 9.2 Nested Samples

We suppose that at some stage in the analysis the input vector $x$ is partitioned as

$$x = \left\{x^s, x^g, x^{p-s-g}\right\}$$

with the understanding that

$$\begin{aligned} x^s &= \text{ subset of size } s \text{ identified as important} \\ x^g &= \text{ subset of size } g \text{ to be tested for importance} \\ x^{p-s-g} &= \text{ remaining inputs.} \end{aligned}$$

| Source of Variation/df | Sum of Squares | $E$(Sum of Squares) |
|---|---|---|
| Total $nk-1$ | $\mathrm{SST} = \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{k} (y_{ij} - \overline{y})^2$ | $k(n-1)\sigma_y^2 + (k-1)\overline{\sigma}_e^2$ |
| Inputs Variables $n-1$ | $\mathrm{SSB} = r \sum\limits_{i=1}^{n} (\overline{y}_i - \overline{y})^2$ | $k(n-1)\mathrm{Var}[E(y \mid x)] + (n-1)\overline{\sigma}_e^2$ |
| Linear fit $p(=1)$ | $\mathrm{SSR} = k\left[\sum\limits_{i=1}^{n} (\overline{y}_i - \overline{y})(x_i - \overline{x})\right]^2 / \sum\limits_{i=1}^{n} (x_i - \overline{x})^2$ | |
| Lack of fit $n-p(=1)-1$ | $\mathrm{SSE} = \mathrm{SSB} - \mathrm{SSR}$ | |
| Simulation Variables $n(k-1)$ | $\mathrm{SSW} = \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{j} (y_{ij} - \overline{y}_i)^2$ | $n(k-1)E(\mathrm{Var}[y \mid x]) = n(k-1)\overline{\sigma}_e^2$ |

$$
\begin{aligned}
\widehat{\eta}_x^2 &= \mathrm{SSB}/\mathrm{SST} \\
\widehat{\eta}_{z.x}^2 &= \mathrm{SSW}/\mathrm{SST} \\
\widehat{\rho}_x^2 &= \mathrm{SSR}/\mathrm{SST}
\end{aligned}
$$

Figure 14. Analysis of variance decomposition for initial sample

The following sample design could be used for this situation. As before, we assume simple random sampling instead of a Latin hypercube sample. Let

$$\{x_t^s,\ t = 1, \cdots, s\}$$

be a random sample form $f_{x^s}$. For each $x_t^s$, let

$$\left\{x_{ti}^g,\ i = 1, \cdots, g\right\}$$

be a simple random sample, independent for each $i$, from $f_{x^g|x^s}$. Next, for each $x_{ti}^g$, let

$$\left\{x_{tij}^{p-s-g},\ j = 1, \cdots, r\right\}$$

be a random sample of the remaining input variables $x^{p-s-g}$, independent for each $j$, from $f_{x^{p-s-g}|x^s,x^g}$ Finally, for each $x_{tij}^{p-s-g}$, let

$$\left\{z_{tijh},\ h = 1, \cdots, k\right\}$$

be a random sample of the simulation variables $z$, independent for each $h$, from $f_{z|x}$. Let the resulting model predictions and computer experiment be

$$\begin{aligned} \{y_{tijh},\ t = 1, \cdots, s; \quad i = 1, \cdots, n \\ j = 1, \cdots, r; \quad h = 1, \cdots, k\} \\ y_{tijh} = m\left(\left\{x_t^s, x_{ti}^g, x_{tij}^{p-s-g}\right\}, z_{tijh}\right) \\ x_i^s \sim f_x \text{ iid} \\ x_{ti}^g \sim f_{x^g|x^s} \text{ conditionally iid} \\ x_{tij}^g \sim f_{x^{p-s-g}|x^s,x^g} \text{ conditionally iid} \\ z_{tijh} \sim f_{z|x} \text{ conditionally iid.} \end{aligned}$$

The general type of products of this step are indicated in Figure 15.

| Source of Variation/df | Sum of Squares | $E$(Sum of Squares) |
|---|---|---|
| Total<br>$rnk - 1$ | $\text{SST} = \sum_{i=1}^{r} \sum_{j=1}^{n} \sum_{h=1}^{k} \left(y_{ijh} - \overline{y}\right)^2$ | — |
| Input subset $x^s$<br>$r - 1$ | $\text{SSB} = nk \sum_{i=1}^{r} \left(\overline{y}_i - \overline{y}\right)^2$ | — |
| Within subset $x^s$<br>$r(nk - 1)$ | $\text{SSWx} = \sum_{t=1}^{s} \sum_{j=1}^{n} \sum_{k=1}^{r} \left(y_{tjk} - \overline{y}_t\right)^2$ | — |
| Input subset $x^{p-s}$<br>$r(n - 1)$ | $\text{SSBw} = k \sum_{i=1}^{r} \sum_{j=1}^{n} \left(\overline{y}_{ij} - \overline{y}_i\right)^2$ | — |
| Simulation Variables<br>$rn(k - 1)$ | $\text{SSW} = \sum_{i=1}^{r} \sum_{j=1}^{n} \sum_{h=1}^{k} \left(y_{ijh} - \overline{y}_{ij}\right)^2$ | — |

$$\begin{aligned} \widehat{\eta}^2_{x^s} &= \text{SSB/SST} \\ \widehat{\eta}^2_{x^{p-s}.x^s} &= \text{SSBw/SST} \\ \widehat{\eta}^2_{z.x} &= \text{SSW/SST} \end{aligned}$$

Figure 15. Analysis of variance decomposition for subsequent samples

Strategies for selecting $s$ inputs in $x^s$ and those in $x^g$ are as numerous as those of variable selection in regression analysis. Identifying important subsets of inputs in uncertainty analysis parallels identifying important subsets in regression analysis. A main difference between the two is that regression analysis is most often carried out under the linear analysis model, whereas we are doing this *importance analysis* under the assumptions of the general analysis model. There is a definite sample trade-off required between the two, as indicated in Figure 16.

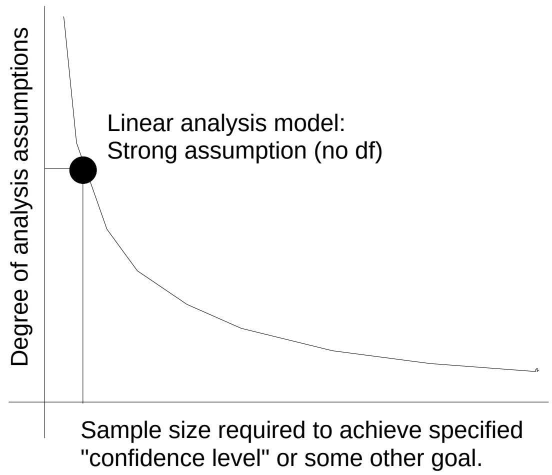


Figure 16. Trade-off between assumptions of analysis and sample size

# 10 Analysis Procedures for the Computer Experiment

The procedures we recommend for uncertainty analysis, those used in the case study of Section 11, follow McKay (1995), in which a complete discussion can be found. There are three objectives: quantification of prediction uncertainty, assessment of contribution to prediction uncertainty from simulation variability relative to that from input uncertainty, and relative importance of individual inputs and input subsets to prediction uncertainty. The extension of the procedures to stochastic simulation models can be made in two ways. If computing resources allow, enough repetitions are obtained at each stage so that the simulation distribution of the outputs are estimated with negligible estimation error. In the more likely event of limitations on computing resources, simulation variables are treated like an unspecified subset of input variables. These procedures are expected to mature as experience with stochastic simulation models is obtained.

## 10.1 Phases of the Analysis

We divide the analysis into three phases: problem definition, screening, and validation. **Problem definition** covers

- Selection of model calculations to use as outputs, predictions, and model observations.
- identification and specification of model input variables, and
- assignment of input probability distributions.

Screening means the identification of input variable subsets indicated as being important. Because demonstrating importance to a high degree of confidence is a computing intensive task with current methodology, it is carried out heuristically in two stages: screening and validation. **Screening** covers

- construction of base case runs where all inputs vary, and
- sequentially fixing inputs and making new runs in anticipation of reducing prediction variance.

In the **validation** phase, possibly several different subsets $x^s$ are constructed from the input variables identified as important in the screening stage. A two-stage experimental design is used in the estimation of appropriate correlation ratios and conditional prediction distributions.

## 10.2 Sequential Estimation and Analysis Steps

The three phases of analysis are composed of many steps, which are summarized as follows.

(1) Identify and describe all potential parameters or input and simulation variables. They fall into three categories: those relating to the numerical algorithms, those describing phenomenology or mechanics of the process being modeled, and those describing the event or scenario being studied.

(2) Identify any inputs which will not be further considered and state how they will be assigned values.

(3) Identify any subsets of inputs that cannot be varied independently.

(4) Choose ranges of variation for those inputs that can be varied independently.

(5) Define domains for each subset of inputs that cannot be varied independently.

(6) In the absence of preferred distributions, assign uniform or loguniform distributions to independent inputs.

(7) Define appropriate joint distribution functions for dependent subsets of inputs.

(8) Obtain base case sample where all inputs vary.

(9) Determine important subsets of inputs:

   (a) Initial stage analysis. Important inputs are determined for each output. Those inputs not in any of the lists are deemed unimportant.

   (b) Subsequent analyses. Separate sequential screening is done for each output to determine important inputs. As before, those inputs not in any of the lists are deemed unimportant.

   Sequential analyses serves several purposes and can produce a more complete subset of important inputs the more the model deviates from linearity.

(10) Perform suitable validation and diagnostic testing:

   (a) Unless a single subset of important inputs for all outputs is to be identified, each model output should be analyzed independently relative to its own subset of important inputs.

   (b) If important subsets do not sufficiently account for prediction uncertainty, continue with the sequential input selection in (9).

   (c) Data from the analysis is examined to reveal any previously undetected relationships and behaviors.

(11) In consultation with subject matter experts, determine final probability distributions for important inputs, and assign the same preliminary distributions to the unimportant ones.

(12) Choose among alternative submodels.

(13) Repeat (8)–(10).

(14) Examine sensitivity of results to perturbation of input distributions. This step is another uncertainty analysis in itself, where the “inputs” define the real input distributions. Whether a formal or informal analysis is carried out is a matter of choice.

(15) Continue if any corrective actions appear necessary.

# 11 Case Study

## 11.1 The AFM Mobility Model

Table II gives examples of input variables from the Airlift Flow Model (AFM).[7] The AFM is the Airlift Simulation portion of the Mobility Analysis Support System (MASS). The input to AFM is a set of interrelated files that describe a cargo movement scenario, the routing structure, the aircraft, the personnel, and the resources available to perform the movement. It is presumed that by changing these resources, the effect of any particular part of the system on the final delivery of the cargo in the scenario can be investigated.

A short description of processes in the AFM would address

- Assignment of cargos to carriers,
- Routing carriers,
- Scheduling of carriers, and
- Miscellaneous tasks.

TABLE II
Classes of input variables in AFM

| |
|---|
| • Time-phased force and deployment data (TPFDD) of airlift movement requirements |
| • An airlift network of on loads, off loads, en route stops, recovery bases, and home stations connected by user-defined routes |
| • An airlift fleet mix of user defined numbers of aircraft types identified by individual tail numbers |
| • Individual aircrews who must be available to allow missions to be flown |
| • Logistics factors which account for refueling, maintenance, and material handling of cargo |
| • Concepts of operation that include strategic airlift, aerial refueling, theater shuttle operations, direct delivery operations, and recovery and staging operations |

In the MASS History Notes[8] we find: “Introduced stochastic event times. In prior releases, certain mission execution events (taxi-out, takeoff, departure, approach, landing, and taxi-in) used constant times. For greater realism in mission execution, uniform distributions are now used for each of the above events to randomly produce mission execution times. By making multiple runs with

[7] Administratively, AFM is one of many modules in the AMC general tool kit called Mobility Analysis Support System or MASS, which is under continuous development and maintenance by HQ AMC/XPY. AFM is the fifteenth such model developed by AMC (and its predecessor, MAC) since the command’s adoption of system simulation in 1975. This document describes Version 4.0 of AFM.

[8] The MASS History Notes exist as a Los Alamos National Laboratory internal Web page, acquired from XPY.

TABLE III
Classes of stochastic variables in AFM

| | |
|---|---|
| • military crew flying hour initialization by aircraft type | • command & control event |
| • en route winds | • resource seizure |
| • weather update | • resource service time |
| • random stats clock | • mission execution approach time |
| • resource prohibition | • mission execution landing time |
| • theater selection for aircraft reallocation | • resource failure |
| • aircraft recheck time | • mission execution taxi-in time |
| • shuttle extension probability | • mission execution taxi-out time |
| • home station check | • resource return to service after failure |
| • aircraft load variability | • mission execution takeoff event |
| | • mission execution departure event |

TABLE IV
Classes of output variables in AFM

- Aircraft measures of effectiveness such as use rates, individual and average payloads, ground service time, flight times, and system delays
- Aircrew measures such as crew duty day, number of crews, hours flown by each crew, and crew availability
- Such cargo status total tons delivered, tons per day throughput, unit and force closure, actual million ton-miles per day flown, cargo remaining in backlog
- Such airlift network statistics as typical cycle times, flying times, network airfield use, MOG constraints, and system bottlenecks

different initial seeds, one can explore the stochastic nature of the airlift system." The times of these events are stochastic variables. Classes of stochastic variables used in the AFM are given in Table III.

For the purpose of evaluating "goodness" of a simulation, various measures are calculated in the AFM. We use these variables as model observations and outputs when studying simulation variability and prediction uncertainty. Classes of output variables are given in Table IV

Finally, a functional view of the MASS model is presented in Table V.

TABLE V
The MASS model

**Mission Planning in the MASS model accomplishes:**

- Prioritized route selection and reservation for flight planning
- Marrying a specific aircraft tail number to the next eligible requirement
- Crew planning to ensure only the crews eligible to fly can fly

**Mission Execution in the MASS model simulates:**

- Events including taxi-out, takeoff, departure, en route cruise, initial approach, final approach, landing, taxi-in, and ground activities for every sortie of every mission
- Ground activities which simulate resource allocation and planned delays for ramp space, off loading cargo, refueling, maintaining, on loading cargo, and crew changing
- Crew activities and monitoring events which include crew rest, crew monthly and quarterly flying hour limits, crew availability, and crew searches for unavailable crews

TABLE VI
AFM output variables for case study

| AFM output description | Output variable names |
|---|---|
| tons delivered by aircraft type | C-141B.t, C-17.t, C-5A.t |
| hours flown by aircraft type | C-141B.h, C-17.h, C-5A.h |

## 11.2 Results from Analysis Phases

*Some version of the AFM was used to simulate some standard exercise. Jack, please fill this in.*

### 11.2.1 Problem Definition

The output variables recorded from each AFM run are given in Table VI. Each output was recorded for $t = 1, 2, \cdots, 15$ days for two random replicates. Random replicates were generated by random sampling, without replacement, from the AFM list of initial random seeds. Eight input variables identified for study and the sampling distributions used in the input uncertainty analysis are provided in Table VII.

TABLE VII
AFM input variables for case study

| Input Variable | AFM File | Type / Range | AFM Input Variable(s) Value |
|---|---|---|---|
| MOG | 25 | Factor<br>U (.3, 1.7) | mog_available<br>Multiplier of variable for all bases with value less than 999. Round result to nearest integer. |
| Max Wait | 31 | Value<br>U (2, 10) | xmax_unif_wait<br>Assign variable for all aircraft types |
| Use Rate | 31 | Factor<br>U (.3, 1.7) | TWO VARIABLES: target_use_rate_delta AND use_rate_delta2<br>Multiply BOTH variables for all aircraft types |
| Enroute Time | 32 | Factor<br>U (.3, 1.7) | std_enroute_time<br>Multiply variable for all aircraft types |
| Offload Time | 32 | Factor<br>U (.3, 1.7) | std_offload_time<br>Multiply variable for all aircraft types |
| Onload Time | 32 | Factor<br>U (.3, 1.7) | std_onload_time<br>Multiply variable for all aircraft types |
| Initial Flying Hours (median) | 47 | Factor<br>U (.3, 1.7) | param 4: alpha<br>Multiply variable for all aircraft types |
| Fuel Flow | 48 | Factor<br>U (.3, 1.7) | fuel_flow<br>Multiply variable for all aircraft types |

### 11.2.2 Sequential Screening

The effect of fixing input variables at their nominal values is observed in stages on the output variables from the AFM. In the base case, the 96 runs are obtained with all variables sampled as described in the next section. In subsequent stages, the sampled values of various subsets of input variables are replaced by the inputs' nominal values; values of the other inputs remain as set for the base case.

**Set 0: Base case sample, no variables fixed**

The 96 base case runs were obtained from $r = 4$ independent replicates of a Latin hypercube sample (McKay, Conover and Beckman, 1979) of size $n = 12$ on the $p = 8$ input variables in Table VII. Each of the $4 \times 12 = 48$ combination of 8 input values (design points) was used in two AFM runs with different random number seeds. The resulting 96 runs for the outputs are displayed in Figure 18. The nominal output values obtained with the inputs each set to their midrange nominal values are displayed for two different random number seeds in Figure 17.

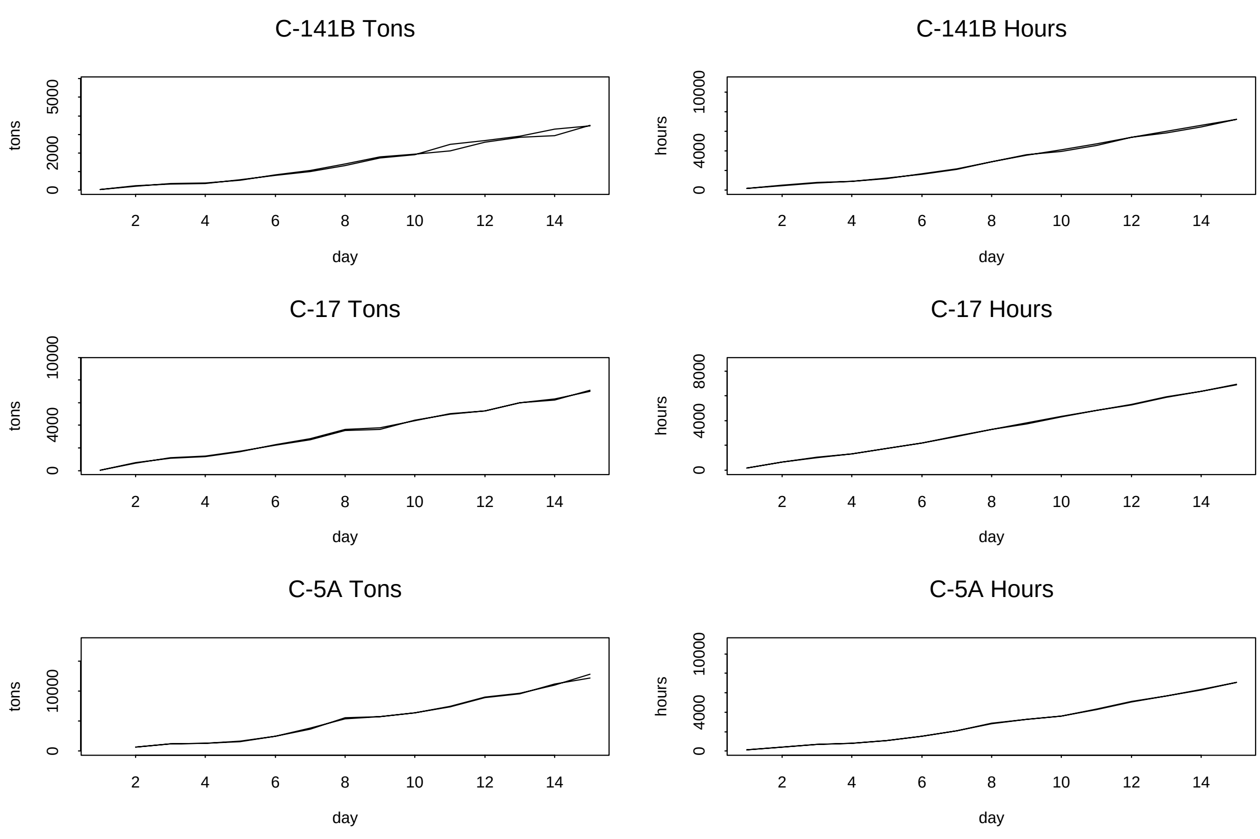


Figure 17. Nominal output values

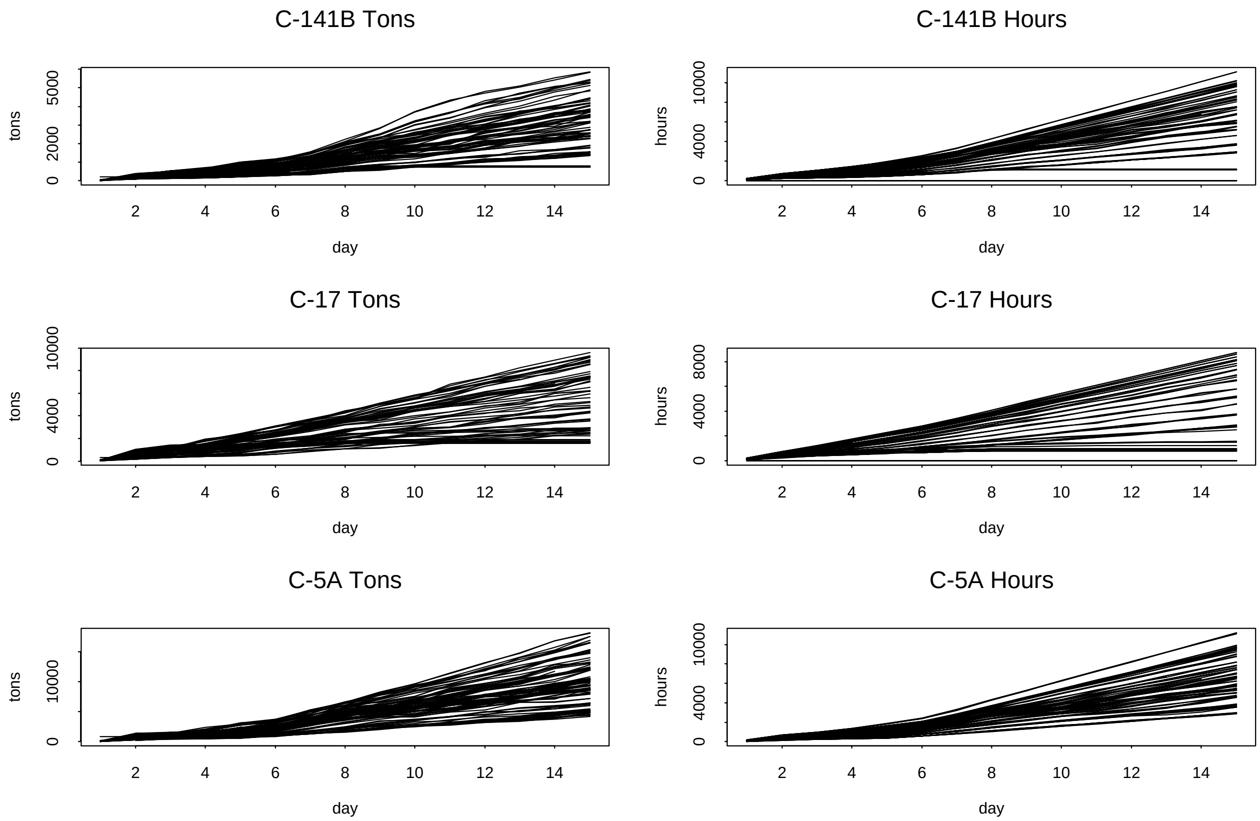


Figure 18. Runs from base case, Set 0

The prediction distribution for each output is to be inferred from the plots in Figure 18. Each of the 6 output variables would have a probability distribution of values at each of 15 days. The same scales are used on the plots for subsequent sets so that reduction in variance as inputs are fixed can be easily observed.

For each output variable, the estimated correlation ratio $R^2$ between its simulation mean value (average over the 2 random replicates) and each of the 8 inputs was computed. Table VIII provides summary statistics for two of the output variables, C-141 Tons and C-141 Hours. The critical value referred to in the table is for guidance only, to help focus attention. The number comes from a test of the hypothesis that $\eta^2 = 0$ under the assumption of a joint normal distribution of $x$ and $y$.

Table VIII is typical of what we found for the 6 outputs. Both of the input variables Use Rate and Fuel Flow are clearly important and dominate the other 6 variables. Use Rate was chosen as the single variable to fix at its nominal value for a rerun of the 96 runs of the AFM. Results are reported in the next section.

TABLE VIII
$R^2$ correlation ratios for base case, Set 0. Critical value $CV = 0.39$

| Output ($\overline{\sigma}$) | | MOG | Max Wait | Use Rate | Enroute Time | Offload Time | Onload Time | Initial Hours | Fuel Flow |
|---|---|---|---|---|---|---|---|---|---|
| C-141.t (1227) | Avg $R^2$ | 0.28 | 0.23 | 0.45 | 0.14 | 0.27 | 0.28 | 0.17 | 0.40 |
| | % days $R^2 \geq CV$ | - | - | 93 | - | 7 | - | - | 47 |
| C-141.h (1782) | Avg $R^2$ | 0.24 | 0.21 | 0.49 | 0.11 | 0.29 | 0.31 | 0.17 | 0.42 |
| | % days $R^2 \geq CV$ | - | - | 100 | - | - | - | - | 53 |

**Set 1: Use Rate fixed**

The 96 input design points from Set 0 with all values of Use Rate set to its nominal value were rerun with the AFM. The observed output values are displayed in Figure 19. The runs show a marked reduction in variability as compared to those in Figure 18. The difference in variability is attributable to Use Rate. The reduction in average standard deviation for C-141.h from 1782 to 1409 is not a true measure of the reduction in its variability. Examination of the runs in Figure 19 shows that the large standard deviation is due to two groups of runs with different means. Much of the remaining variability is explainable by Fuel Flow, as anticipated from $R^2$ from Set 0 and as indicated in Table IX, which is, again, typical of what we found for the 6 outputs.

**Set 2: Use Rate and Fuel Flow fixed**

The 96 input design points from Set 0 with all values of Use Rate and Fuel Flow set to their nominal value were rerun with the AFM. The observed output values are displayed in Figure 20. The runs show a continuing reduction in variability as compared to those in Figures 18 and 19. The reduction in variability is attributable to Use Rate and Fuel Flow. Together, these two inputs reduce the variability in the outputs by 90%, as measured by "average" standard deviation: the square root of the average over the 15 days of the squared daily standard deviation from the 96 runs,

$$\overline{\sigma} = \sqrt{\frac{1}{15}\sum_{d=1}^{15}\frac{1}{95}\sum_{i=1}^{96}\left(y_{id} - \overline{y}_d\right)^2} .$$

At this point in the analysis, the input MOG is dominant for the Tons output variables, and the input Enroute time is dominant for the Hours outputs. The $R^2$ from Set 2, indicated in Table X are typical of what we found for the 6 outputs.

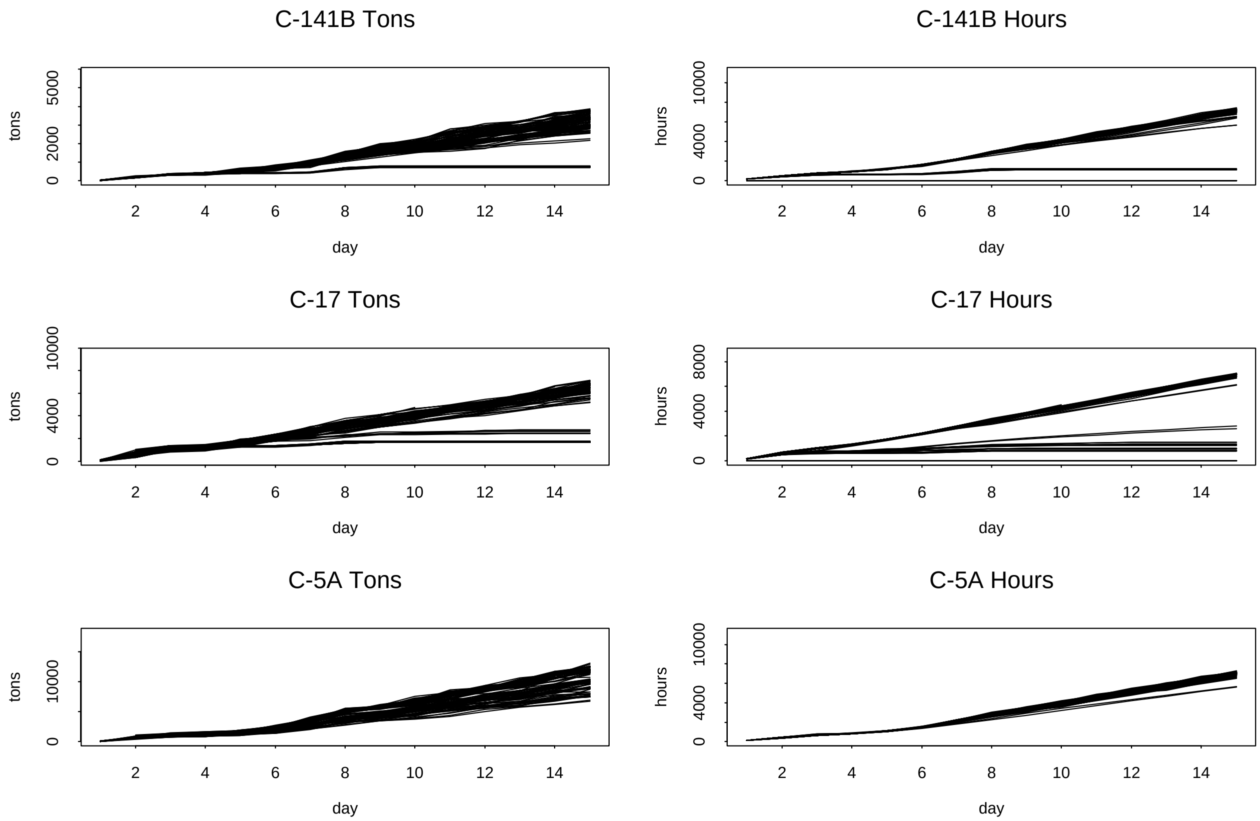


Figure 19. Runs from Set 1, Use Rate set to nominal value.

TABLE IX
$R^2$ correlation ratios for Set 1. Critical value $CV = 0.39$

| Output ($\overline{\sigma}$) | | MOG | Max Wait | Use Rate | Enroute Time | Offload Time | Onload Time | Initial Hours | Fuel Flow |
|---|---|---|---|---|---|---|---|---|---|
| C-141.t (646) | Avg $R^2$ | 0.34 | 0.37 | Fixed | 0.29 | 0.23 | 0.26 | 0.14 | 0.72 |
| | % days $R^2 \geq CV$ | 13 | 40 | - | 7 | - | 7 | - | 100 |
| C-141.h (1409) | Avg $R^2$ | 0.25 | 0.26 | Fixed | 0.29 | 0.22 | 0.35 | 0.20 | 0.69 |
| | % days $R^2 \geq CV$ | - | - | - | 7 | - | 40 | - | 100 |

**Set 3: Use Rate, Fuel Flow and Enroute Time fixed**

The 96 input design points from Set 0 with all values of Use Rate, Fuel Flow and Enroute Time set to their nominal value were rerun with the AFM. The observed output values are displayed in

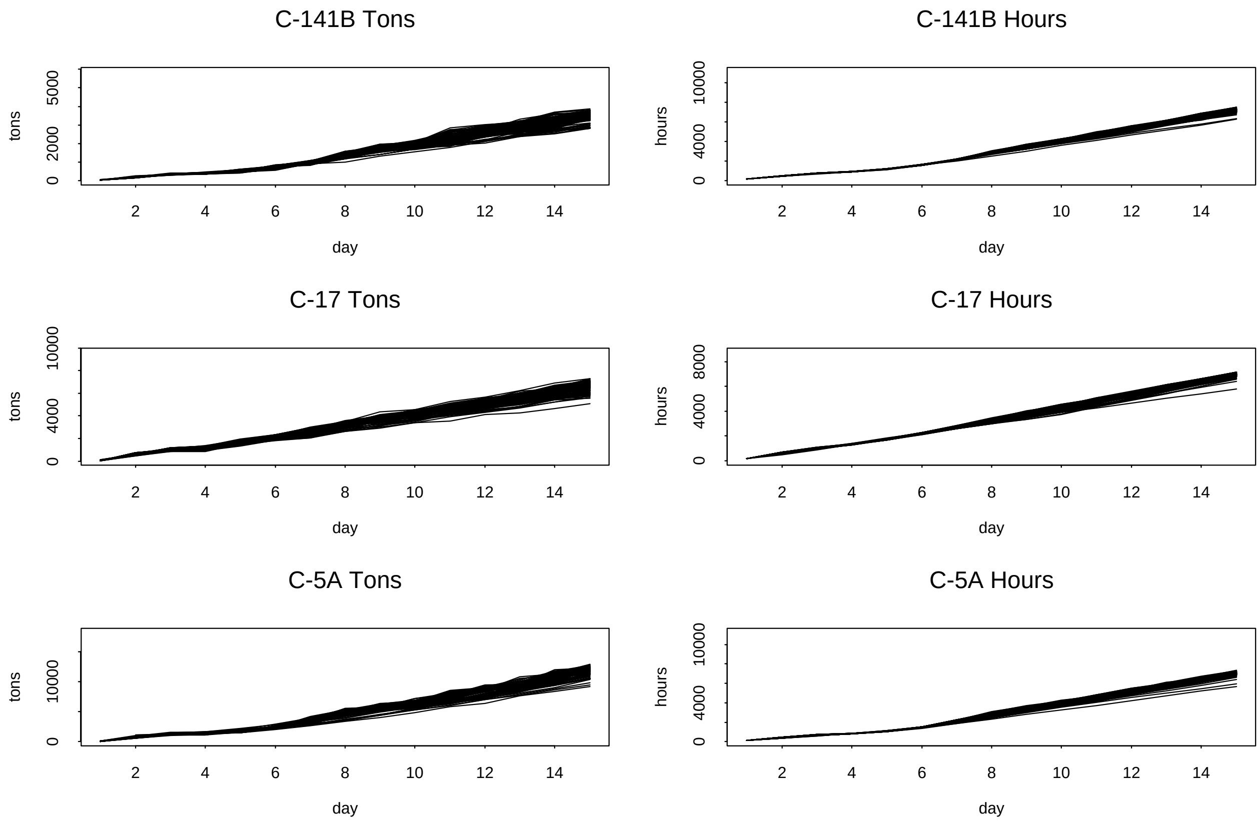


Figure 20. Runs from Set 2, Use Rate and Fuel Flow set to nominal values.

TABLE X
$R^2$ correlation ratios for Set 2. Critical value $CV = 0.39$

| Output ($\overline{\sigma}$) | | MOG | Max Wait | Use Rate | Enroute Time | Offload Time | Onload Time | Initial Hours | Fuel Flow |
|---|---|---|---|---|---|---|---|---|---|
| C-141.t (143) | Avg $R^2$ | 0.52 | 0.40 | Fixed | 0.44 | 0.20 | 0.31 | 0.22 | Fixed |
| | % days $R^2 \geq CV$ | 87 | 53 | - | 23 | 23 | 23 | 23 | - |
| C-141.h (109) | Avg $R^2$ | 0.39 | 0.23 | Fixed | 0.54 | 0.30 | 0.39 | 0.16 | Fixed |
| | % days $R^2 \geq CV$ | 53 | 20 | - | 67 | 13 | 20 | - | - |

Figure 21. The runs show a reduction in variability in the Tons output variables when compared to those in Figure 20, attributable to the input Enroute Time. The reductions are not as dramatic those in previous sets. From a practical perspective, the observed variability in the Hours outputs

is of no consequence. Therefore, we conclude from the screening stage that the inputs Use Rate, Fuel Flow and Enroute Time control all of the essential variability in the three outputs, C-141B.h, C-17.h and C-5A.h.

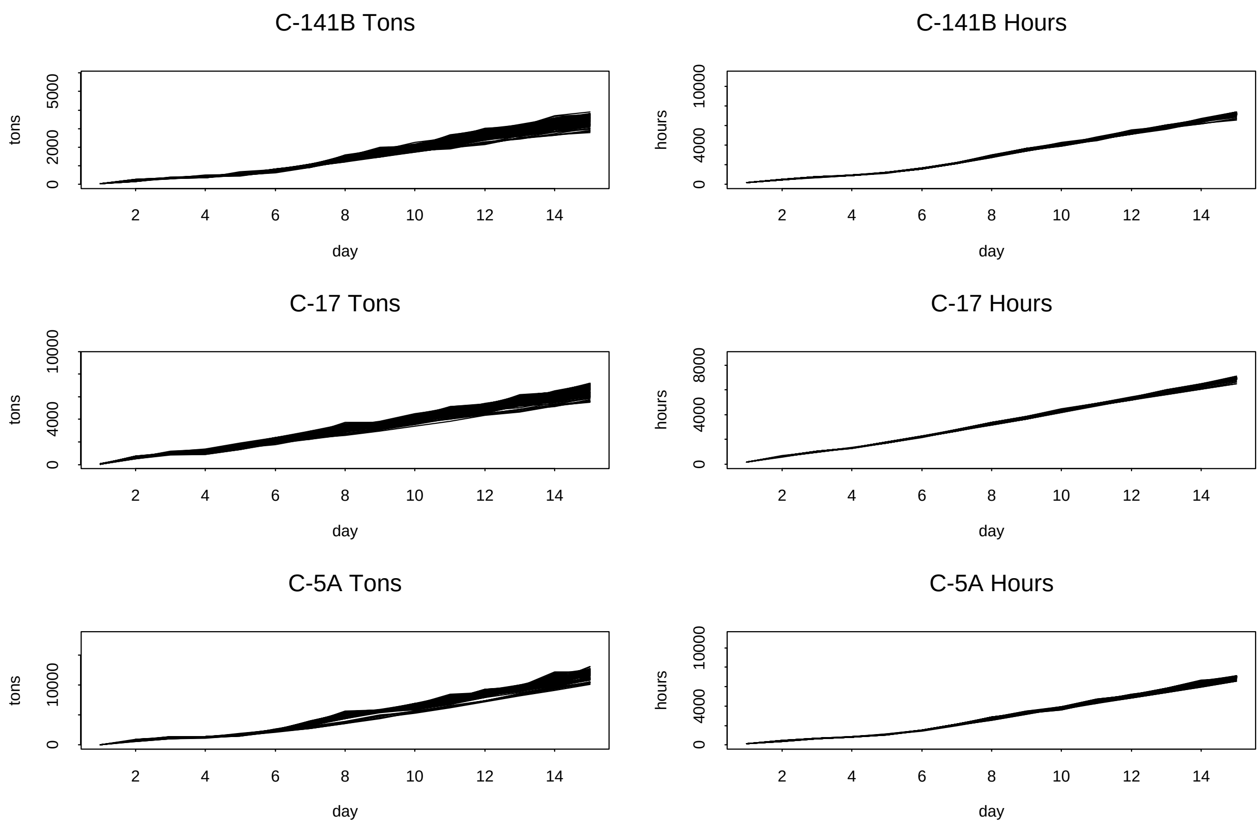


Figure 21. Runs from Set 3, Use Rate, Fuel Flow and Enroute Time set to nominal values.

TABLE XI
$R^2$ correlation ratios for Set 3. Critical value $CV = 0.39$

| Output ($\overline{\sigma}$) | | MOG | Max Wait | Use Rate | Enroute Time | Offload Time | Onload Time | Initial Hours | Fuel Flow |
|---|---|---|---|---|---|---|---|---|---|
| C-141.t (132) | Avg $R^2$ | 0.64 | 0.27 | Fixed | Fixed | 0.26 | 0.36 | 0.18 | Fixed |
| | % days $R^2 \geq CV$ | 87 | 7 | - | - | 13 | 23 | 23 | - |
| C-141.h (57) | Avg $R^2$ | 0.51 | 0.29 | Fixed | Fixed | 0.25 | 0.32 | 0.23 | Fixed |
| | % days $R^2 \geq CV$ | 60 | 7 | - | - | 7 | 20 | - | - |

**Set 4: Use Rate, Fuel Flow and MOG fixed**

The 96 input design points from Set 0 with all values of Use Rate, Fuel Flow and MOG set to their nominal value were rerun with the AFM. The observed output values are displayed in Figure 22. The runs show a reduction in variability in the Hours output variables when compared to those in Figure 20, attributable to the input MOG. The reductions are not as dramatic those in previous sets. From a practical perspective, the observed variability in the Tons outputs is of no consequence. Therefore, we conclude from the screening stage that the inputs Use Rate, Fuel Flow and MOG control all of the essential variability in the three outputs, C-141B.t, C-17.t and C-5A.t.

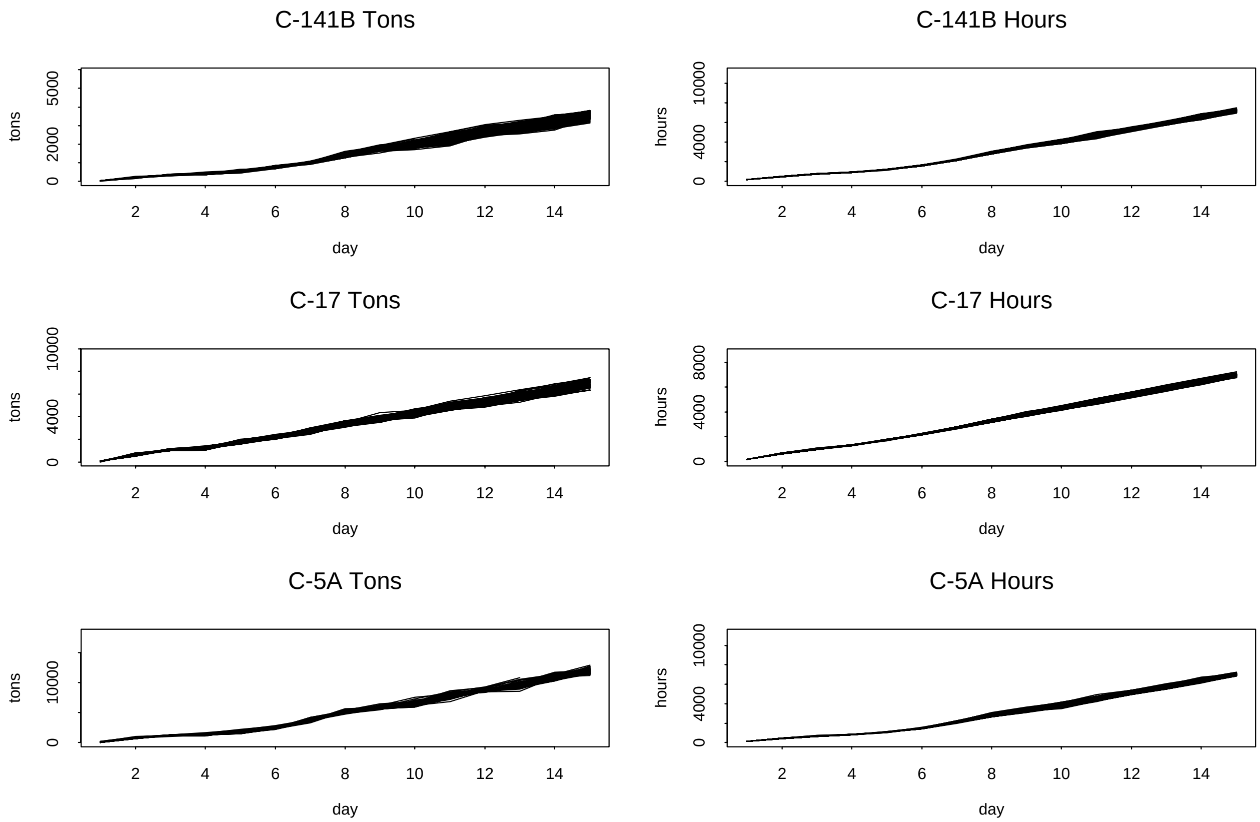


Figure 22. Runs from Set 4, Use Rate, Fuel Flow and MOG set to nominal values.

### 11.2.3 Validation

In the screening stage, input variables selected as important were fixed at their nominal value. Their selection needs to be validated over the entire input space and not just at the nominal. Figure 23 shows the results of fixing the values of the inputs Use Rate, Fuel Flow, MOG and Enroute Time at 8 different points (sets of values) and rerunning the AFM. These runs are like those for Set 4 but with 8 different values for the “nominal” settings. The runs demonstrate how the 4 important inputs change the general patterns of the outputs of AFM and, at the same time, the residual variability associated with the remaining 4 inputs.

TABLE XII
$R^2$ correlation ratios for Set 4. Critical value $CV = 0.39$

| Output ($\overline{\sigma}$) | | MOG | Max Wait | Use Rate | Enroute Time | Offload Time | Onload Time | Initial Hours | Fuel Flow |
|---|---|---|---|---|---|---|---|---|---|
| C-141.t (108) | Avg $R^2$ | Fixed | 0.25 | Fixed | 0.48 | 0.51 | 0.24 | 0.25 | Fixed |
| | % days $R^2 \geq CV$ | - | 23 | - | 23 | 23 | 23 | 23 | - |
| C-141.h (81) | Avg $R^2$ | Fixed | 0.25 | Fixed | 0.67 | 0.33 | 0.27 | 0.16 | Fixed |
| | % days $R^2 \geq CV$ | - | 7 | - | 93 | 47 | 7 | - | - |

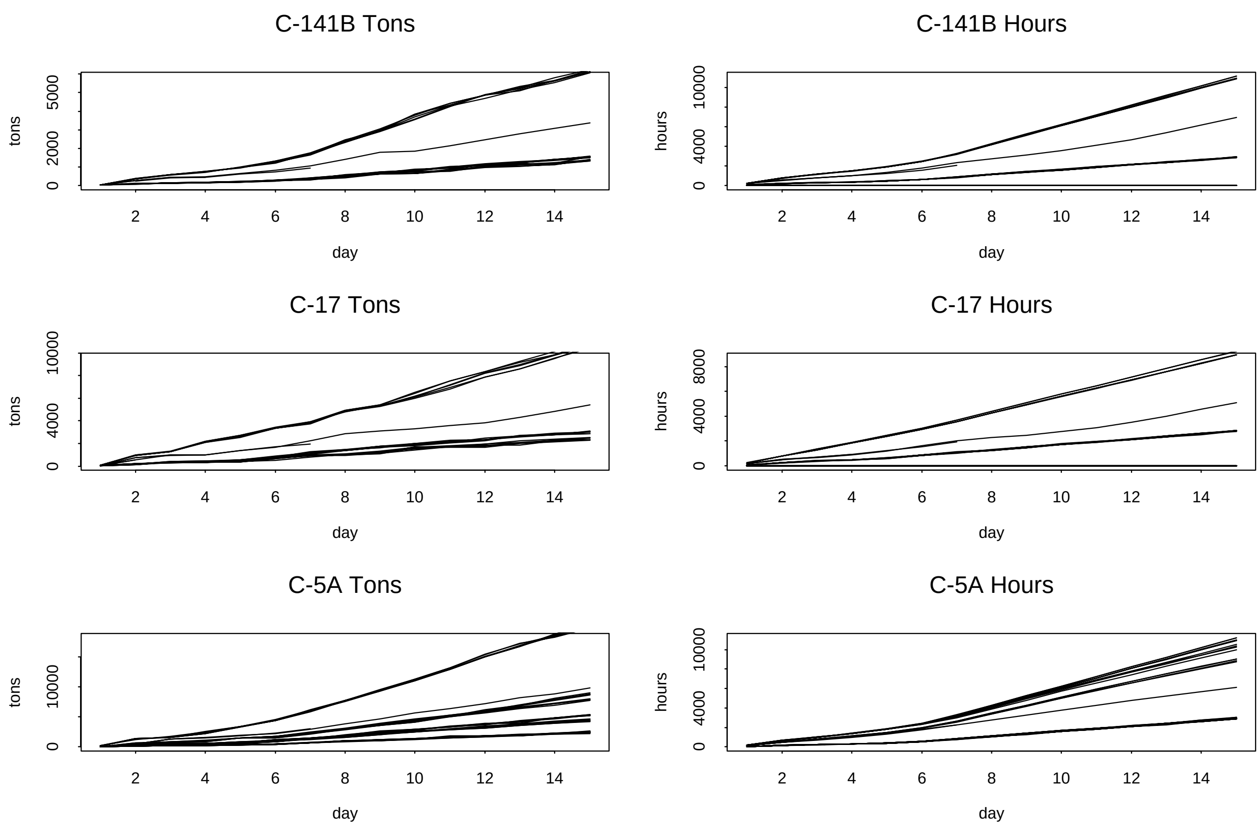


Figure 23. Validation runs with important inputs Use Rate, Fuel Flow, Enroute Time and MOG set to 8 combinations values.

In Figure 24, the 4 unimportant inputs are set to their nominal values while the 4 important inputs vary. The ranges of observed values closely parallels those in Figure 18, indicating that the important inputs, alone, can produce essentially all of the variability in the outputs.

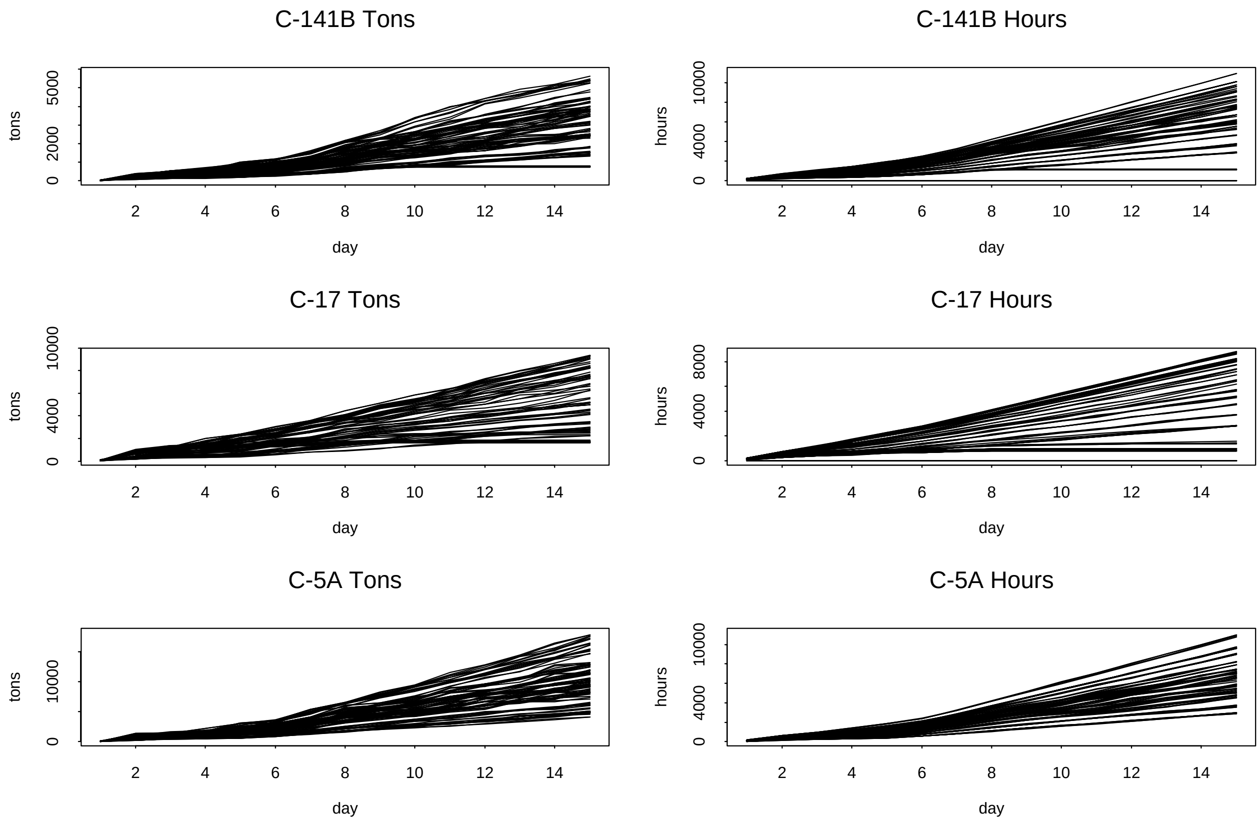


Figure 24. Validation runs with unimportant inputs Max Wait, Offload Time, Onload Time and Initial Hours set to their nominal values

## 12 Mobility Models — A Reality Check

Mobility models are described as in Figure 25. The figure indicates three modules or components of a mobility model: Assigner, Scheduler, Simulator. Requirements, Assets, Carriers, and Cargos are lists, each item of which has attributes. Assigner and Scheduler (Planner) are processes. They might be rule based, optimizing or hybrid algorithms. They might be combined into a single algorithm, as indicated by the dashed box. The Simulator (Engine) is a stochastic compartmental model. The circles represents lists whose elements are collections of attributes. The boxes broadly function to associate elements from their input lists, modify values of the attributes, and produce output lists. The model operates by repeated cycling through the three modules and an update module. For this study of model uncertainty, we group the assignment and scheduling processes together in the Planner, and confine attention on the Simulator. The function of the Planner is to construct a plan as an ordered sequence of activities or operations to be executed by the Simulator. At execution, the Simulator determines the feasibility of the plan and associates times with activities.

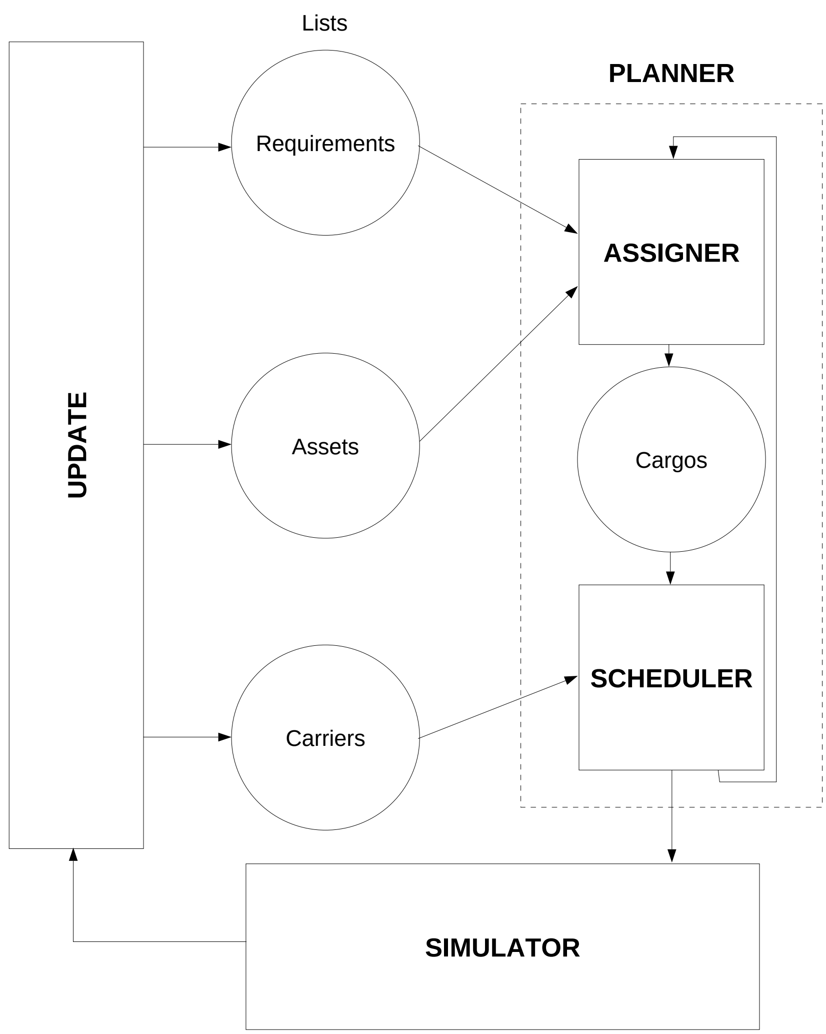


Figure 25. Class of admissible mobility models

The *prediction* of a mobility model to consist of two parts: the sequence of activities and their start and duration times. The quality of the prediction, therefore, depends on differences between predicted and observed activity sequences, and on differences between predicted and observed times.

Metaphorically speaking, we assume each module, and in particular the Scheduler, comprises descriptions of activities which "model" or move each object/actor from activity to activity until the actions necessary for it are completed, at which time the actor exits the module. In this manner, a model is a directed graph where each vertex (node) represents not necessarily unique activities, and every path from an initial node to a final node represents an admissible transmission through the model. We define a *residence time* $r$ at each node, which may be a function of the system rules and the system state. The system state is defined to be a specification of the position and residence times of each actor in their graph at time $t$. Thus, we have the definition of a model in the class mobility model. Resolution of the model has to do with the fine structure of the graph. In a coarsely resolved system, a vertex and its resident time comes from a very crude approximation of an activity. As the activity is better defined from constituent parts and times, finer, less aggregated models evolve.

We consider an example for the actor Aircraft. We suppose that each aircraft is one of a type of Carriers, for which attributes are defined. Activities (vertices, nodes) and compartmental models determine residence times (to complete the activity and depending of carrier type) are defined as follows.

- At Home — null state, no constraint
- Traveling — $r = f(\text{origin, destination, load, preformance})$
- On loading — $r =$ standard time
- At En-route Stop — $r =$ standard time
- Aerial Refueling — $r =$ standard time
- Off loading — $r =$ standard time
- Recovery — $r =$ standard time

The Simulator models a *mission* for aircraft $i$ as a sequence of activities,[9] denoted by

$$
\begin{gathered}
A_{ij} = j^{\text{th}} \text{ activity for aircraft } i \\
j = 1, 2, \cdots, n_i \\
i = 1, 2, \cdots, I
\end{gathered}
$$

We denote the set of possible activities for $A_{ij}$ by

$$A_{ij} \in \{\alpha_k \mid k = 1, 2, \cdots, n_\alpha\} .$$

[9] The notion of server location is included in the activity. It may be treated separately.

Associated with each activity is a residence time, $r_{ij}$, determined by the compartmental model associated with the activity. That is,

$$A_{ij} = \alpha_k \text{ for some } k, \text{ begun at time } t_{ij}$$
$$r_{ij}(t_{ij}) = r(\alpha_k, S(t_{ij})),$$

where

$$t_i = t \text{ for actor } i$$
$$S(t) = \text{ state of system at } t\ .$$

From the perspective of a queuing model, the residence time is the sum of the waiting (for service) time and the service time.

Activities for an actor are depicted in Figure 26. The figure illustrates issues related to the quality of the model prediction, namely, the selection and sequencing of activities $\alpha_k$, their positioning $t_i$ on the event time line, and their durations or residence times $r_i$. The questions we ask about the predictions are:

- Are there missing or extra activities?
- Are the activities in the correct sequence?
- What is the uncertainty in activity event times ($t_i$)?
- What is the uncertainty in the activity residence times ($r_i$)?

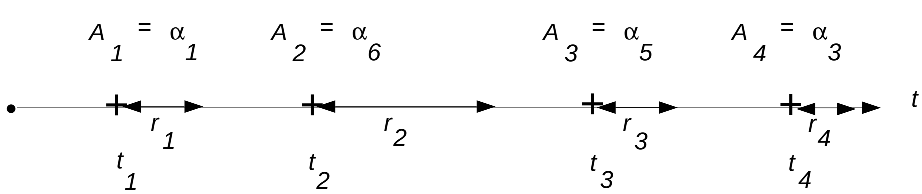


Figure 26. Actor's view: time line of 4 activities

### 12.1 Compartmental Model for Activities

The Simulator in Figure 25 is a stochastic simulation model which can be represented as a collection of stochastic compartmental models representing the activities associated with aircraft moving cargos. Level of detail and aggregation define the substructure of a compartment. For a low level of detail (resolution), the single dashed compartment is indicated in Figure 27. For more detail, the single compartment is replaced by the six interior ones. In the highest level, the inner compartment in the upper right is replaced by six other compartments. An actor (aircraft) remains in an activity compartment from a residence time which depends on the characteristics of the activity. A context for analysis of structural uncertainty using compartmental models of activities and residence time is presented in the next section.

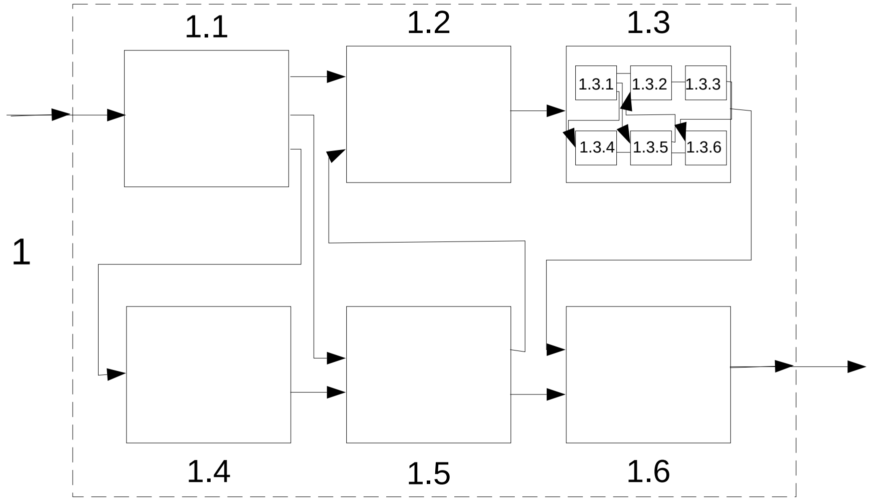


Figure 27. Refinements of compartmental models

# 13 Context for Structural Model Uncertainty

We do not address a general notion of model, but a particular one of the discrete event simulation model and, even more specifically, a mobility model. The approach taken is to design a prototype or super model from an abstraction of a mobility model. From such a prototype a strategy for analysis of model uncertainty would be constructed. This section present only a beginning to the approach.

Practicality and usefulness of concepts and measures related to structural uncertainty depend on context. A reasonable way to characterize structural model uncertainty is to pattern it after input uncertainty. Input uncertainty is reflected in the prediction distribution $f_y$ induced by a probability distribution $f_x$ on inputs $x$ for fixed model parameters $\theta$ and fixed model $m(\cdot)$. The context for assessing model input uncertainty, then, is

$$\{x \in D, f_x, \theta;\ y = m(x;\theta)\}\ .$$

The input domain $D$ is usually reasonably well known. Model parameter values are often not well known and, as a result, $\theta$ is often treated like a model input with an uncertainty distribution $f_\theta$. The principal weakness of the context for input uncertainty is in justification the functional form of $f_x$.[10] The prediction distribution and, in particular, the prediction variance are induced by the input distribution $f_x$, which is why the form of $f_x$ is important. Justification for $f_x$ notwithstanding, meaningful and useful assessment of model input uncertainty can be accomplished with proper caveats.

Contexts for assessing structural model uncertainty analogous to the one for assessing model input uncertainty arise in a natural way. A description of such a context, also found in Draper (1995), is

$$\{m \in \mathcal{M}, g_m, x, \theta;\ y = m(x;\theta)\}\ , \tag{13.1}$$

where specification of inputs and parameters is assumed to depend on the model $m(\cdot)$. There are several significant problems with this context, however. First of all, except in limited cases, the model domain $\mathcal{M}$ exists at most inferentially and, usually, only hypothetically. That is, while the possibility of alternative models is not excluded, their existence is only presumed. As a result, $y$ may only be observable for a limited number (like, one) of models. Secondly, and more importantly, both the meaning and form of $g_m$ are virtually unknown. The meaning of $g_m$ is an open question. Two possibilities are (1) $g_m$ is a sampling distribution on $\mathcal{M}$ (which seems unlikely) and (2) $g_m$ represents the likelihood (objective or subjective) that any particular model is correct with respect to reality (a seemingly equally unlikely possibility). The mathematical construct of a relevant family $\mathcal{M}$ of models is, nevertheless, necessary though much more difficult to come by than, for example, the specification a domain of input values.

[10] Sensitivity of results, namely, estimates of $f_y$, to different forms of $f_x$ are prudently evaluated using, for example, Beckman and McKay (1987).

## 13.1 Family of Models

We formally denote the family $\mathcal{M}$ by

$$\begin{aligned} \mathcal{M} = &\text{ a family of simulation models} \\ &= \left\{ m_\psi,\ \psi \in \Psi \right\} . \end{aligned} \tag{13.2}$$

From this point, there are several directions that might be explored in the search for methods to assess the uncertainty in model predictions arising from choice of model in $\mathcal{M}$. A very general approach for discrete event simulation models would use the context of Eq. 13.1 with arbitrary notions of state and state transition. The mechanism for describing the evolution of the system over time would be arbitrary. From such a general context, nothing more that formal definitions of prediction uncertainty arising from structural uncertainty seem to be available.

Another direction for exploration in search of methodologies for analysis of effects of structural uncertainty on model prediction examines not models in general but specific types or classes of models. This is the approach taken in this document. The generic model considered is expressible in some form as a logic diagram. Event-tree models fall into this class. Some questions of structural uncertainty relate to the degree of detail with which a real system in the world is modeled in, say, the structure of an event tree. Other questions of structural uncertainty relate to accuracy of processes for determining probabilities of state transitions and movement through the tree. Processes modeling to determine transition probabilities depends on both representation of the real system and laws of nature.

The essence of an event-tree model is the specification of possible sequences of events together with their probabilities of occurrences, so that the probability of a top event can be calculated. We choose to define the system being modeled as a collection of *actors* together with a collection of *processes*. The actors are entities transformed from state to state in activities defined by processes. Therefore, a system description comprises descriptions of all of the actors. The diagrammatic structure of event trees might describe sequences of states for each actor and/or states of the system as a whole. Importantly, process are defined through computational units and identifiable as modules or submodels. Notions of state and state transitions are restricted, from an actor's point of view, to those expressible with logic diagrams, like directed graphs. In summary:

- A model consists of actors and processes.
- Dynamics of the world are represented by the processes which transform the actors from state to state through activities.
- The modeled system evolves over times as a result of the action of the processes.

Under this notion of model, one aspect of structural uncertainty concerns detail of representations:

- detail of representation of actors with respect to the world,
- detail of representation of processes with respect to the world, and
- detail of representation of states and activities with respect to the world.

The modeler may not free to treat these facets independently because they impose requirements and restrictions of each other. Therefore, their effects on prediction uncertainty may not be completely separable in an analysis of structural uncertainty. So, without trying to be specific at this point, we suppose that a model $m(\cdot)$ is written as

$$m(x) = m_h(x, h(x)) ,$$

where $h(\cdot)$ symbolizes an identifiable module or submodel within $m(\cdot)$. The members of $\mathcal{M}$ are denoted by

$$m_\psi(x) = m_h\big(x, h_\psi(x)\big) ,$$

indicating alternative choices for the module $h(\cdot)$. The function $h(x)$ can be a vector of submodels or a single submodel or module within the model of $m(\cdot)$. The functions $\{h_\psi(x);\ \psi \in \Psi\}$ would be alternative formulations or replacements of $h(x)$. If $m_\psi\big(x, h_\psi(x)\big) = h_\psi(x)$ then the representation is equivalent to the general one of Eq. 13.2. Therefore, without loss of generality, we reduce the problem of assessing structural model uncertainty to the problem of assessing structural *submodel* uncertainty. That is, we are able to break into pieces the question of structural model uncertainty and focus attention to smaller portions[11] by two limitations on context: model type (event tree) and submodel replacement.

## 13.2 Prediction Error and Structural Uncertainty

We propose that the specific structure of $m(\cdot)$ be used to propagate the structural uncertainty in any individual module $h_\psi$ to uncertainty in prediction. In particular, we want to discern how changes (improvements) in a module are propagated as reductions in uncertainty in prediction and in prediction error. When $h_\psi$ is complex and vector valued, it may be necessary to propagate changes on a component-by-component basis. The propagation might be accomplished in three ways:

- analytically, using an assumed logical structural for the model;
- numerically, as with input uncertainty and using submodel substitutions or transformations of their computed values; and
- approximately, using a response surface approximation or a first order analysis of variance model like

$$y_{ijkuv} = h_{\psi_i} + \theta_{j|i} + x_{k|ij} + (\text{solution algorithm})_{u|ijk} + (\text{simulation variability})_{v|ijku} .$$

[11] It may turn out that the submodel are smaller, in some sense, but just as complex and difficult to work with.

The terms in the analysis of variance model correspond to sources of variation: submodel structure $h$, parameter values $\theta$, input values $x$, possibly approximate solution algorithm, and random number stream.

Ideally, the quality of a model prediction would be evaluated with a metric on the prediction error, denoted

$$\|y - w\| ,$$

where $y = m(x;\theta)$ is the model prediction for input values $x$ and parameter values $\theta$, and $w$ is the corresponding observation of reality. A common choice for the metric is the mean-squared error of prediction

$$\begin{aligned} \Delta^2 &= E(y-w)^2 \\ &= (\mu_y - \mu_w)^2 + \left(\sigma_y^2 - 2\rho_{yw}\sigma_y\sigma_w + \sigma_w^2\right) . \end{aligned} \tag{13.3}$$

The expectation integral $E$ is with respect to a joint probability distribution of reality $w$ and model prediction $y$. Although it is not clear what that probability distribution could be, it would arise from input and parameter uncertainty, conditional on the model $m(\cdot)$, together with a probability distribution describing uncertainty in reality. The distribution would determine the means $\mu$, standard deviations $\sigma$ and correlation coefficient $\rho$. Eq. 13.3 illustrates the bias, $(\mu_y - \mu_w)$, and precision components, $\left(\sigma_y^2 - 2\rho_{yw}\sigma_y\sigma_w + \sigma_w^2\right)$, of the mean-squared error of prediction. When the model prediction $y$ is perfectly correlated with observation $w$ (which is not likely), then: $\rho_{yw} = 1$, the precision component is minimum, and the mean-squared error of prediction becomes

$$\Delta^2 = E(y-w)^2 \rightarrow (\mu_y - \mu_w)^2 + (\sigma_y - \sigma_w)^2 .$$

Thus, mean-squared error of prediction is zero under the conditions[12] that the model is

- perfectly correlated with reality, $\rho_{yw} = 1$,
- correctly centered, $\mu_y = \mu_w$, and
- properly scaled, $\sigma_y = \sigma_w$.

Although the three parameters of correlation, location and scale which determine the quality of a submodel might be used to compare submodels, it is not possible to judge the values of the parameters in an absolute sense without information on $w$. Of the three, the most important and most difficult to assess is the correlation $\rho_{yw}$. Effects of differing location and scale parameters are more easily investigated. Effective methods for the joint assessment of the three parameters are needed.

Limitations on available data $w$ almost always make evaluation of prediction error impossible. Therefore, “expert opinions” are used in substitution for, or to supplement, measured information on $w$. This form of information might lead to the diagram in Figure 28, where an approximate relation between the location and scale parameters of the mean-squared error of prediction are indicated. Therefore, assessment of improvement is likely to be more assessment of changes in value (rather than in absolute magnitudes) with assumptions about direction and distance for improvement.

[12] Values of the parameters change with scenario and model.

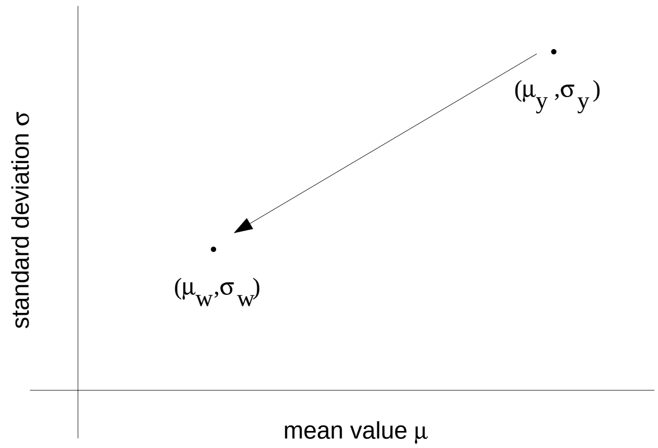


Figure 28. Two components of mean-squared error of prediction for model prediction $y$ and reality $w$.

### 13.3 Refinement and Disaggregation

We assume that, up to a point, the mean-squared error of prediction $\Delta^2$ can be controlled by the level of detail used in a submodel $h$. For mobility models, increase in detail in $h$ is related to refinement of activity (process) models and to decrease in aggregation of actors with corresponding refinement of the definition of system state. We also assume that more detail in the module $h$ provides a better representation of reality which, in turn, would make $y$ closer to $w$.

Let $h(x)$ have the counterpart $\eta$ in nature. We assume that $y$ is improved by making $h(x)$ closer to $\eta$. We can think of $h(x)$ as predicting $\eta$, and apply the metric of mean-squared error of prediction to the prediction error $h(x) - \eta$ in the same way as it is applied to the model prediction. However, there are still difficulties in observing $\eta$ just like those in observing $w$. Thus, we think of improving $h(x)$ by adjusting its mean, its standard deviation, and its correlation with $\eta$.

For mobility simulation models, we are concerned with event times for specified activities. Therefore, we denote the prediction of the module $h$ by $t$, and consider a sequence of refinements, or alternative formulations, as

$$t_k = h_k(x)$$

with

$$E(t_k \mid \eta) = \mu_k = \eta + b_k$$
$$\text{Var}(t_k \mid \eta) = \sigma_k^2 \,.$$

Appealing to Figure 28, we look at a sequence of improvements to $h$ as one where the sequence of biases $b_k$ converges[13] to zero—a strong assumption—and that the sequence of conditional variances

[13] Convergence is used in the sense of convergence in probability.

$\sigma_k^2$ also converges to zero—another strong assumption. When $\lim_{k\to\infty} b_k = 0$ and $\lim_{k\to\infty} \sigma_k^2 = 0$, and when the correlation between submodel prediction and reality approaches 1, then

$$\begin{aligned}\lim_{k\to\infty} \text{Var}(t_k) &= \lim_{k\to\infty} \{E(\text{Var}(t_k \mid \eta)) + \text{Var}(E(t_k \mid \eta))\} \\ &= \lim_{k\to\infty} E(\sigma_k^2) + \lim_{k\to\infty} \text{Var}(\eta + b_k) \\ &= \sigma_\eta^2 \,.\end{aligned}$$

Following the perturbation method of McKay (1993), the event times $t$ determined by $h(x)$ might be transformed (biased) in an assumed direction of $(\mu_\eta, \sigma_\eta^2)$ to provide a quantitative assessment of variability due to process uncertainty in $h(\cdot)$.

Assessment of structural uncertainty associated with the detail of representation of actors seems more difficult. At the highest level of detail, an actor in the model corresponds to the smallest item acted upon in reality. At a low level of detail, an actor represents an ensemble of items in reality, for which the process models operate on their behavior and characteristics. Therefore, assessment of effect of aggregation of actors requires that substantial consideration be given to the process modeling. A consideration for mobility models is taken up in the next section.

## 13.4 Activity Models

General concepts related to structural model uncertainty coming from leaving the notion of model output general provide formal descriptions that seem to have little practical benefits. Therefore, we appeal to the descriptions of Section 12 to look to specifics of mobility modeling and queuing models, in particular, as providing a foundation upon which to develop practical uncertainty analysis. We assume that structural model uncertainty is restricted to uncertainty in activity times but not in the sequence of activities for each actor. An activity $A_i$, modeled by a submodel $h$, has several times associated with it:

$$\begin{aligned} t_i &= \text{ arrival time for } A_i \\ q_i &= \text{ waiting time for } A_i \\ s_i &= \text{ service time for } A_i \\ r_i = q_i + s_i &= \text{ residence time for } A_i \end{aligned} \tag{13.4}$$

$$\begin{aligned} t_i &= t_{i-1} + r_{i-1} \\ t_0 &= q_0 = r_0 = 0 \,. \end{aligned}$$

Although the waiting and services times are distinct within an activity model, it is sufficient for this discussion to consider only their sum, which we call the “residence time.” Various forms of queuing models might be used to determine waiting and service times. These activity models would reflect resource limitations, and, hence, also determine delay times. For simplicity, delay times are assumed to be system-determined rather than process- or activity-determined. That is, once an actor begins an activity—gets into an activity compartment—the activity takes place without interaction with the rest of the system. After completion of the activity, the actor may be delayed because of system state.

Simple activity models $h$ might use multiple queues and multiple servers with exponential waiting and service times, for example, which seems to be the approach used in the AFM. Refinement of a model $h$, as mentioned before, is taken on in two dimensions: the detail of representation of the actor and the detail of representation of the activity or process. However, the output of the model remains a single value, the residence time. The objective of model refinement is to improve the prediction of residence time.

We let a refinement $\alpha_k^\dagger$ of activity $\alpha_k$ be defined by a partition of the activity into sequential subactivities. That is,

$$\alpha_k^\dagger = \alpha_{k1} \to \alpha_{k2} \to \alpha_{k3} \to \cdots \to \alpha_{kq} \ ,$$

where the residence time of the refinement is additive[14] in those of the subactivities,

$$R^\dagger\left(\alpha_k^\dagger\right) = \sum_{j=1}^{q} R(\alpha_{kj}) \ .$$

By refining the activity, we assume the quality[15] of the modeled residence time improves over that of the aggregated model of the activity. Therefore, we want to relate changes in method of residence time determination to changes in a performance measure of the model as a whole.

### 13.5 Example of Refinement of Activity Residence Times

The aspect of model refinement important to us is how it relates to improvement of prediction of residence time, where improvement is measured relative to nature. Suppose, as before, that an activity $\alpha$ is refined to $k$ subactivities. Let

$$\begin{aligned} r &= \text{ residence time under } \alpha \\ r^\dagger &= \text{ residence time under the refinement } \alpha^\dagger \end{aligned}$$

For a simple base case, suppose that the $\alpha$–model for $r$ is

$$r \sim \text{Exponential}(\mu_r)$$

with $\mu_r$ taken to be a constant representing the scenario average residence time. Suppose that the refined $\alpha^\dagger$–model uses covariates $x$ to produce the residence time:

$$r^\dagger = r \mid x \sim \text{Exponential}\left(\mu_r^\dagger(x)\right)$$

where for example, a linear regression might be used for the mean function,

$$\mu_r^\dagger(x) = x\beta \ .$$

Refinement occurs when information $x$ is used to improve the model for $r$. But how does one know that $r^\dagger$ is a real improvement of $r$? Or that the overall model prediction $y$ is improved? Usually, the effect of changing to $r^\dagger$ can only be assessed qualitatively, through a shift in $(\mu_y, \sigma_y)$. However, it is reasonable to assume that propagation of the changes in an activity model can be assessed more quantitatively using the model structure. Research in that direction is recommended.

[14] There is a distinction between series and parallel circuits, particularly because their residence times add differently. The times probably add like resistances in electric circuits.

[15] Quality will eventually have to be defined, probably in the mean-squared error sense.

## 14 Conclusions

A general mathematical foundation for uncertainty analysis of stochastic simulation models is presented. The foundation provides a reasonable and effective basis to relate prediction uncertainty, input uncertainty and simulation variability through decomposition of prediction variance. The value of correlation ratios, under the general analysis model, is demonstrated in the case study.

The methods of this paper could be applied to transformations of the model output. Two common transformations are ranks and logarithms. A study of the properties resulting from using these transformations, including interpretation of importance indicators, although not performed as part of this research, would be of benefit to the analysis community.

This document presents directions for research and development of methods for assessing effects of structural model uncertainty on model prediction based, mostly, on suppositions. The document's principal conclusion is that viable generic methods for assessing structural model uncertainty do not now exist nor is their development eminent. It is recommended that research be continued from a practical perspective wherein a context for structural uncertainty be developed for and applied to an existing model, like the Air Flow Model. Not only would such an approach allow for testing and evaluation of the theory and conjectures contained in this document, but more importantly, its hands-on style would likely provide insight and guidance for formulating first principals and a first-principals approach to assessing structural model uncertainty.

Along the lines of the development in this document, we propose that attention first would be limited to the Simulator within the AFM. To perform an analysis of the AFM, its generic description in Section 12 would be expanded to identify and describe all components or submodels that are candidates for replacement from some family of submodels $\mathcal{M}$. Having so defined activities and processes within the AFM, their sequential dependencies and interdependencies would be indicated in process logic diagrams. The process diagram—see, for example, Figure 27—would be used to construct activity time lines—see, for example, Figure 26—for each actor in the simulation. The process diagrams and time lines would form a basis for constructing a (complete) AFM logic diagram for which a propagation of variance would be used in the evaluation of structural uncertainty.

It was argued in Section 13 that AFM-like simulation models may be studied by looking at three activity times: arrival, waiting, and service. Replacement of submodels determining the times can be looked at as the process of activity refinement discussed in Section 13.4. Therefore, we propose that a schematic description relating times, activities, and processes be constructed for the (complete) AFM.

Direct replacement of modules in the AFM would allow investigation of the structural model uncertainty for the process for a fixed level of aggregation of the actors. Aggregation in the AFM takes place in two dimensions: process and actor. For processes, each may be viewed from microscopic to macroscopic levels. For actors, each might represent a microscopic or distinct individual in reality, or a macroscopic or ensemble average.

Structural model uncertainty really addresses two issues: how well a model uses inputs $x$ to determine the output $y$, and how well $x$ alone can, in the limit, predict $y$. Generally, process modeling and refinement relate to the first point, and notions of aggregation and ensemble averaging in $x$ relate to the second. It is a common assumption that the principal way to improve quality of prediction in discrete event simulation models is to increase detail of representation of system states. While we do not argue the point, we propose that both detail of state representation and detail of process modeling—determining state transition probabilities—be investigated for their effects on predictions for the AFM. An anticipated by-product of the investigation, besides gaining knowledge of models in the class of AFM, is valuable ideas on how one might construct a general framework to investigate and assess in structural model uncertainty the roles and contributions of level of detail in process modeling and state representation.

If it is true for AFM that submodel uncertainty can be studied by looking at activity times like those in Eq. 13.4, then appropriate schematic representations and logic diagrams for the AFM would be constructed to be used to relate the times to a measure of performance. We expect that a completely analytical propagation of variance will be difficult to develop. Therefore, we would develop numerical methods to bias the times and produce a variance decomposition similar to those used in analysis of input uncertainty. At this time, we think an idea that holds promise is that of constructing a super model to generate times, based on the ideas of Sacks, et al., (1989), and using variables and interdependencies indicated by the logic diagrams of the AFM.

It is not apparent how one might construct a complete, valid mathematical abstraction of a model as complex as the AFM. However, such an abstraction would likely use concepts from Markov processes, graph theory, dynamical systems, an so forth. We believe that the undertaking of the theoretical approach to studying structural model uncertainty of the AFM should follow and be based on a pragmatic and heuristic study. Therefore, any theoretical work should be at a lower priority and be done to suggest directions and to verify findings for the AFM application study.

## Appendix: Notation

- $m(\cdot)$: a model. The triple $(\Omega, T, \mathbf{P})$
- $\omega \in \Omega$: states and state space.
- $t \in T$: time and set of time points.
- $\mathbf{P} : \Omega \to \Omega$: transition probability function, $\mathbf{P} = \{p_{ij}\}$.
- $\theta \in \Theta$: model parameters and parameter space.
- $\mathbf{P}_\theta$: from a parametric family of transition probability functions, $\{\mathbf{P}_\theta,\ \theta \in \Theta\}$.
- $\mathcal{M}$: family of models, $\{m_\psi,\ \psi \in \Psi\}$
- $h(\cdot)$: a submodel. $m(x) = m_h(x, h(x))$.

- $\Gamma : \Omega \to \mathrm{R}^d$: observation function on the state space.
- $W$: state in the world and the object of model prediction. $W$ is a random variable defined on the state space $\Omega$.
- $w = \Gamma(W)$: an observation on a world state.
- $w^*, W^*, \omega^* \in \Omega^*$: ultimate descriptions in the world. In context, $w, W, \omega \in \Omega$ would represent partial descriptions.

- $Y$: model predicted state, corresponding to $W$. $Y$ is a random variable defined on the state space $\Omega$. The existence of $Y$ may be only implied.
- $y = \Gamma(Y)$: model output, prediction, an observation on a model-predicted state.

- $\mu_w, \mu_{w|x}$: mean value of $w$, unconditional and conditional on inputs $x$.
- $\mu_y, \mu_{y|x}$: mean value of $y$, unconditional and conditional on inputs $x$.
  *Conditioning on a model is to be understood.*
- $\tau$: the object of prediction of the model output $y$. $\tau$ might be an expected value, a variance, a distribution function, whatever.
- $\tilde{y}$: an estimator, a model mean, for example, $\tilde{y} = \overline{y} = \widehat{\tau} = \widehat{E(w)}$.

# References


Apostolakis, G. (1990). The concept of probability in safety assessments of technological systems. *Science*, 250:1559–1564.

Atwood, C. L. (1993). Individual model evaluation and probabilistic weighting of models. In *Proceedings of Workshop I in Advanced Topics in Risk and Reliability Analysis, Model Uncertainty: Its Characterization and Quantification*, NUREG/CP–0138, pages 99–106, Annapolis, MD. U.S. Nuclear Regulatory Commission.

Bailey, N. T. J. (1964). *The Elements of Stochastic Processes*. John Wiley & Sons, Inc.

Beckman, R. J. and McKay, M. D. (1987). Monte carlo estimation under different distributions using the same simulation. *Technometrics*, 29(2):153–160.

Bratley, P., Fox, B. L., and Schrage, L. E. (1987). *A Guide to Simulation*. Springer-Verlag, second edition.

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society, B*, 57(1):45–97.

Helton, J. C. (1993). Uncertainty and sensitivity analysis techniques for use in the performance assessment for radioactive waste disposal. *Reliability Engineering and System Safety*, 42:327–367.

Helton, J. C. (1994). Treatment of uncertainty in performance assessments of complex systems. *Risk Analysis*, 14(4):483–511.

Iman, R. L. and Hora, S. C. (1990). A robust measure of uncertainty importance for use in fault tree system analysis. *Risk Analysis*, 10(3):401–406.

Kendall, M. and Stuart, A. (1979). *The Advanced Theory of Statistics*, volume 2, chapter 26. MacMillan Publishing Co., New York, fourth edition.

Kotz, S. and Johnson, N., editors (1982). *Encyclopedia of Statistical Sciences*. John Wiley & Sons, New York.

Krzykacz, B. (1990). Samos: A computer program for the derivation of empirical sensitivity measures of results from large computer models. Technical Report GRS-A-1700, Gesellschaft fur Reaktorsicherheit (GRS) mbH, Garching, Republic of Germany.

Kullback, S. (1968). *Information Theory and Statistics*. Dover Publications, New York.

Laskey, K. B. (1996). Model uncertainty: theory and practical implications. *IEEE Transactions on Systems, Man, and Cybernetics*, 26(3):340–348.

McKay, M. D. (1993). Aspects of modeling uncertainty and prediction. In *Proceedings of Workshop I in Advanced Topics in Risk and Reliability Analysis, Model Uncertainty: Its Characterization and Quantification*, NUREG/CP–0138, pages 51–64, Annapolis, MD. U.S. Nuclear Regulatory Commission.

McKay, M. D. (1995). Evaluating prediction uncertainty. Technical Report NUREG/CR–6311, U.S. Nuclear Regulatory Commission and Los Alamos National Laboratory.

McKay, M. D., Conover, W. J., and Beckman, R. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245.

Pearson, K. (1903). Mathematical contributions to the theory of evolution. *Proceedings of the Royal Society of London*, 71:288–313.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science*, 4(4):409–435.

Saltelli, A., Andres, T. H., and Homma, T. (1993). Sensitivity analysis of model output: An investigation of new techniques. *Computational Statistics & Data Analysis*, 15:211–238.

Winkler, R. L. (1993). Modeling uncertainty: Probabilities for models? In *Proceedings of Workshop I in Advanced Topics in Risk and Reliability Analysis, Model Uncertainty: Its Characterization and Quantification*, NUREG/CP–0138, pages 107–116, Annapolis, MD. U.S. Nuclear Regulatory Commission.

Zio, E. and Apostolakis, G. (1996). Two methods for the structured assessment of model uncertainty by experts in performance assessments of radioactive-waste repositories. *Reliability Engineering and System Safety*, 54:225–241.